REPORT OF THE GEOSTATISTICAL ANALYSIS OF HIGH RESOLUTION MULITSPECTRAL IMAGERY

Final Report (RSSUSA - 3/4)

Dr Margaret A Oliver and Professor Richard Webster
U of Reading. UK
February 1998 - May 1998

United States Army

ENVIRONMENTAL RESEARCH OFFICE OF THE U.S. ARMY

London, England

CONTRACT NUMBER - N 68171 - 97 - C - 9029

Contractor - Approved for Public Release; distribution unlimited

REPORT OF THE GEOSTATISTICAL ANALYSIS OF HIGH RESOLUTION MULITSPECTRAL IMAGERY

Final Report (RSSUSA - 3/4)

Dr Margaret A Oliver and Professor Richard Webster

February 1998 - May 1998

United States Army

ENVIRONMENTAL RESEARCH OFFICE OF THE U.S. ARMY

London, England

CONTRACT NUMBER - N 68171 - 97 - C - 9029

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	01.07.98	Final - Feb 1998 - May 1998

4. TITLE AND SUBTITLE
Geostatistical analysis of high resolution multispectral imagery

5. FUNDING NUMBERS
N68171
97-C-9029

6. AUTHOR(S)
Dr Margaret A Oliver, Professor Richard Webster

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Reading, Whiteknights,
Reading, RG6 6DW, UK

8. PERFORMING ORGANIZATION REPORT NUMBER
RSS USA-3/4

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USARDSG-UK, Environmental Sciences Branch
Edison House, 233 Old Marylebone Road,
London, NW1 5TH, UK

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Final Report: Geostatistical analysis of high resolution multispectral imagery

12a. DISTRIBUTION/AVAILABILITY STATEMENT
No limitation on distribution/availability

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This is the final report of the project to apply geostatistics to a high resolution remote imagery. The area chosen for analysis is part of Fort A P Hill in north-eastern Virginia. This report includes some of the earlier reports, the aide memoir for geostatistical analysis, and the latest analysis of the 1m resolution imagery. Part of this project has been used to set up some geostatistical computing capability for TEC. Dr Oliver visited TEC for this in 1996. The initial analysis was of a SPOT image. The area chosen for study, to the north of Anderson Camp was divided into a working area and a validation one. A variogram analysis of the SPOT image revealed two scales of spatial variation. Kriging analysis was performed on the three wavebands and NDVI, and the report contains the maps of the area from the analysis. They show clearly that there are at least two spatial scales of variation that we can identify from the SPOT image. The analysis of the ground cover data from three transects show that three distinct spatial scales emerge: about 200m , 500m to 600m, and 1200m, which accord with those from the multivariate variograms. These spatial scales in ground cover match those in the image data closely. The variogram analysis of the high resolution imagery has shown that although more spatial scales have been identified than from the SPOT image they correspond closely and there is little additional information.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Fort A P Hill, ground cover, high resolution imagery, kriging analysis, aide memoir, spatial scales, SPOT imagery.	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

TABLE OF FIGURES (continued)

TABLE OF FIGURES (continued)

PREFACE

This report was sponsored by the United States Army Research, Development and Standardization Group-UK (USARDGSG-UK), London, NW1 5TH, United Kingdom, under contract PR-N 68171-97-C-9029, entitled, "Geostatistical Analysis of High Resolution Multispectral Imagery - Phase 3", and is monitored by the U. S. Army Topographic Engineering Center (TEC), Alexandria, Virginia 22315-3863. The work was done by the University of Reading, Whiteknights, Reading, RG6 6DW, United Kingdom. The USARDGSG-UK Program Manager is Mr. Jerry Comati, and the TEC Contracting Officer's Representative is Mr. Kevin Slocum.

EXECUTIVE SUMMARY

This report contains most of the earlier interim reports with the latest analysis of the 1 m imagery for Fort A. P. Hill. The aim of this work has been to examine the spatial structures present in part of a SPOT image and in 1 m resolution imagery for the same area at Fort A. P. Hill. The first part of the work on this project involved a visit by Dr Oliver to TEC to help with setting up some computing capability in geostatistics using Genstat and to go into the field to choose study sites. The variogram analysis of the three wavebands and the normalized difference vegetation index (NDVI) of the SPOT data for the study site showed two clear scales of spatial variation: one of about 160 m and the other of about 640 m. The image was then filtered using kriging analysis which uses the different spatial scales identified in turn and kriges it. This filters each scale from the others. The short-range component seems to relate closely to the local variation in relief, and the long-range one to the main blocks of ground cover, such as woodland. Ground cover surveys were designed taking into account the spatial scales identified. The results confirmed a strong relation between the scales of variation in the image data and those in the ground cover.

The 1 m imagery results in enormous files that are too large to edit even on large workstations. We have had to develop a means of handling these. There were four wavebands in the 1 m data. Variograms were computed for each at several sample spacings and for different parts of the study area. The spatial scales identified correspond closely with those from the SPOT data. A short scale of 20 m to 30 m possibly represents shadows from the tree crowns. Other significant scales of variation appear to be 100 m to 200 m, 500 m to 700 m and about 1 km. It seems at this stage of the work that little additional information on significant spatial scales has been obtained from the 1 m imagery. The SPOT image appears to resolve the variation at a sensible spatial scale for management on the ground. Clearly the two different resolutions have different purposes, but for defining the main patterns of variation in ground cover the SPOT data are adequate.

The last section of the report contains an 'aide memoir' for geostatistical analysis and also a suite of programs for Genstat. This provides TEC with the capability to do part of the analysis embraced in the report.

PART I: THE SITE AT FORT A. P. HILL

This project has been based on remotely sensed data from a part of Fort A. P. Hill in northeastern Virginia, about 47 km from Washington, DC. It occupies 30 757 hectares of gently undulating terrain on the Atlantic Coastal Plain of the United States with an average elevation of about 15 m. The area is intensely dissected by many small waterways, some of which drain into ponds, and some of which are surrounded by wet land. Drainage is to two main basins: the Rappahhanock and the James. Hence, there is a significant watershed crossing the Fort. The area is largely forested with mixed woodland, together with some open areas of shrub and grassland. The Fort is much used for training, and the effects of this on the environment are of concern.

One of the aims of this project has been to investigate relations between SPOT imagery, which has a 20 m $\times$ 20 m resolution, and the ground information. This has been supported by a further investigation using high resolution imagery to 1 m.

While Dr Oliver was at TEC in November 1996, two regions were chosen initially for investigation (report Appendix I). One was an area known as the Wildlife Refuge (around 9524), and the other was around Anderson Camp (9616). Our initial analysis was for the former. The variogram of the red waveband of the SPOT image was computed and modelled in preparation for the field the next day. During the field visit the landscape and vegetation of both areas were examined.

The variogram for the Wildlife Refuge showed two distinct spatial scales of variation. On walking through the forest we tried to relate the change in ground cover with the results of the variogram analysis. The shorter range of variation was about 200 m, and this structure was evident, especially in the small patches of evergreen woodland among the predominantly deciduous forest. The latter appears to represent the longer range component of the variation of about 900 m. There were also small open patches that would also reflect the short-range component of the variation. There is also considerable variation in the undergrowth and the relief.

It was decided after the initial work on the Wildlife Refuge to focus on the area to the North of Anderson Camp. One reason for the change is that flights are not allowed over the Wildlife Refuge and the other that the larger area around Anderson camp will provide the scope to use a part of it for validation of any predictions in the future. The study area was chosen during Dr Oliver's second visit to A. P Hill at the end of March 1997. It is a rectangle with the southwestern corner at 970155, 940155 and the northeastern one at 970190, 940190. It has been divided so that we

shall have access to ground information for one half, and the other will provide a basis for validating the information derived from both the SPOT image and the high resolution imagery. The ultimate aim is to assess whether it is possible to predict ground conditions reliably from imagery alone. The ability to do this is vital where ground is held by an enemy. It is also valuable in other areas of potential danger, such as where mines have been laid, and more generally for remote regions and others areas to which access is difficult.

The pixel maps of the three wave bands are shown in Figures 1 to 3. An attempt was made to show approximately the same amount of detail in the variation in each of the wavebands, hence the scales of the reflectance values are different. The NDVI (D), was computed from bands 1 (Red) and 3 (Infra-red) by

$$D = (I - R)/(I + R) ,$$

where I and R are the values in the near infra-red and red channels respectively. The NDVI is also shown as a map, Figure 4.

Table 1: Summary Statistics for SPOT data, A. P. Hill.

Channel	Minimum	Maximum	Mean	Variance	St. Dev.	Skewness
1 (Red)	52.0	127.0	63.09	29.26	5.41	2.49
2 (Green)	27.0	124.0	37.56	50.17	7.08	2.48
3 (NIR)	37.0	183.0	118.15	279.46	16.72	-0.76
NDVI	−0.212	1.214	0.564	0.0451	0.2123	−0.5344

The statistical distributions of the three wavebands and NDVI were determined and are given in Table 1. The data in channels 1 and 2 are fairly strongly positively skewed, but those in channel 3 are approximately symmetrical. We transformed the data in channels 1 and 2 to stabilize their variances using logarithms as we had done for the Fort Benning data, but this was not successful. Several other methods of transformation were tried and there was little improvement, and the analyses were carried out on the raw data.

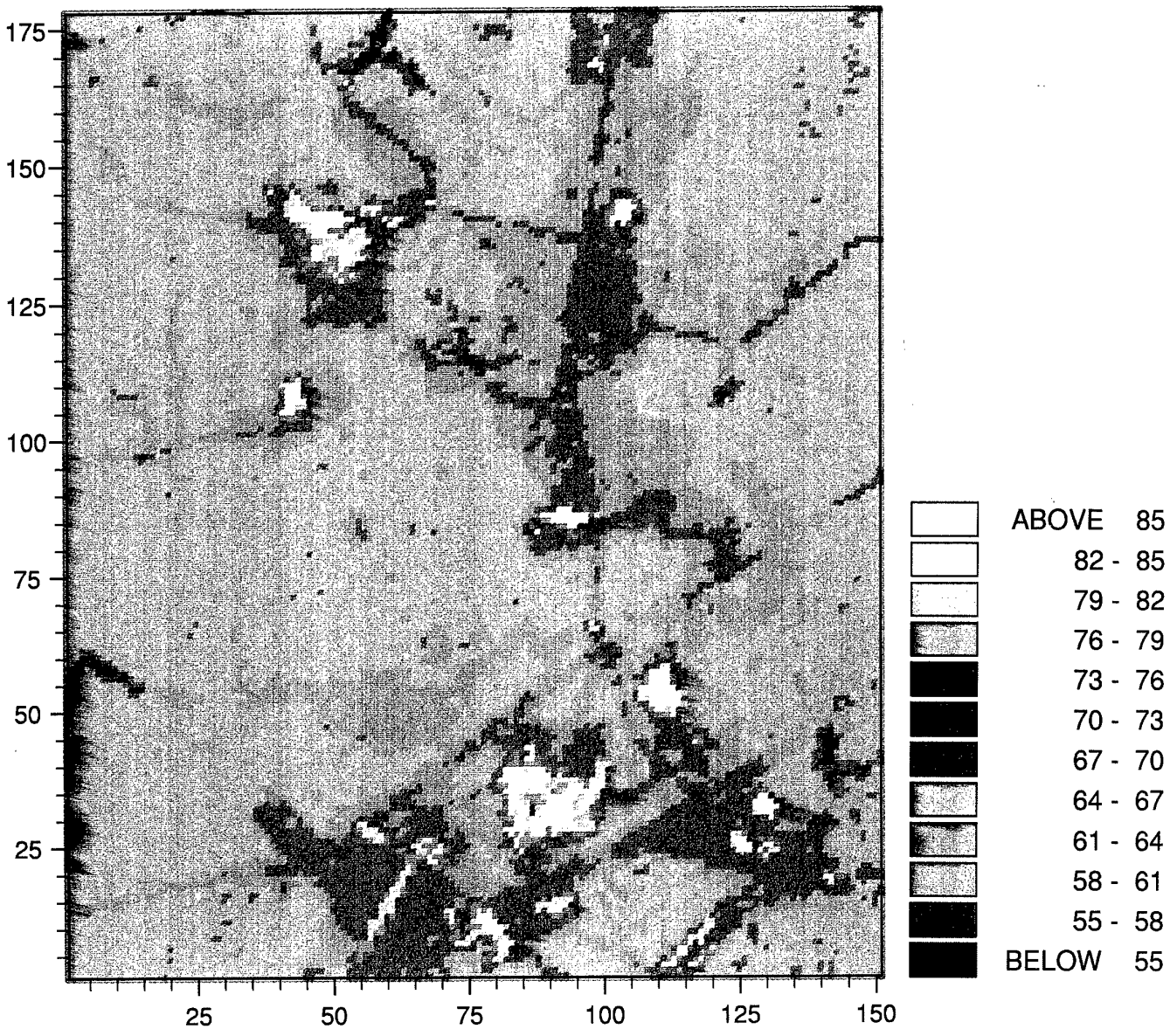

Figure 1: Pixel map of spectral values for channel 1 (Red) from SPOT data, A. P. Hill.

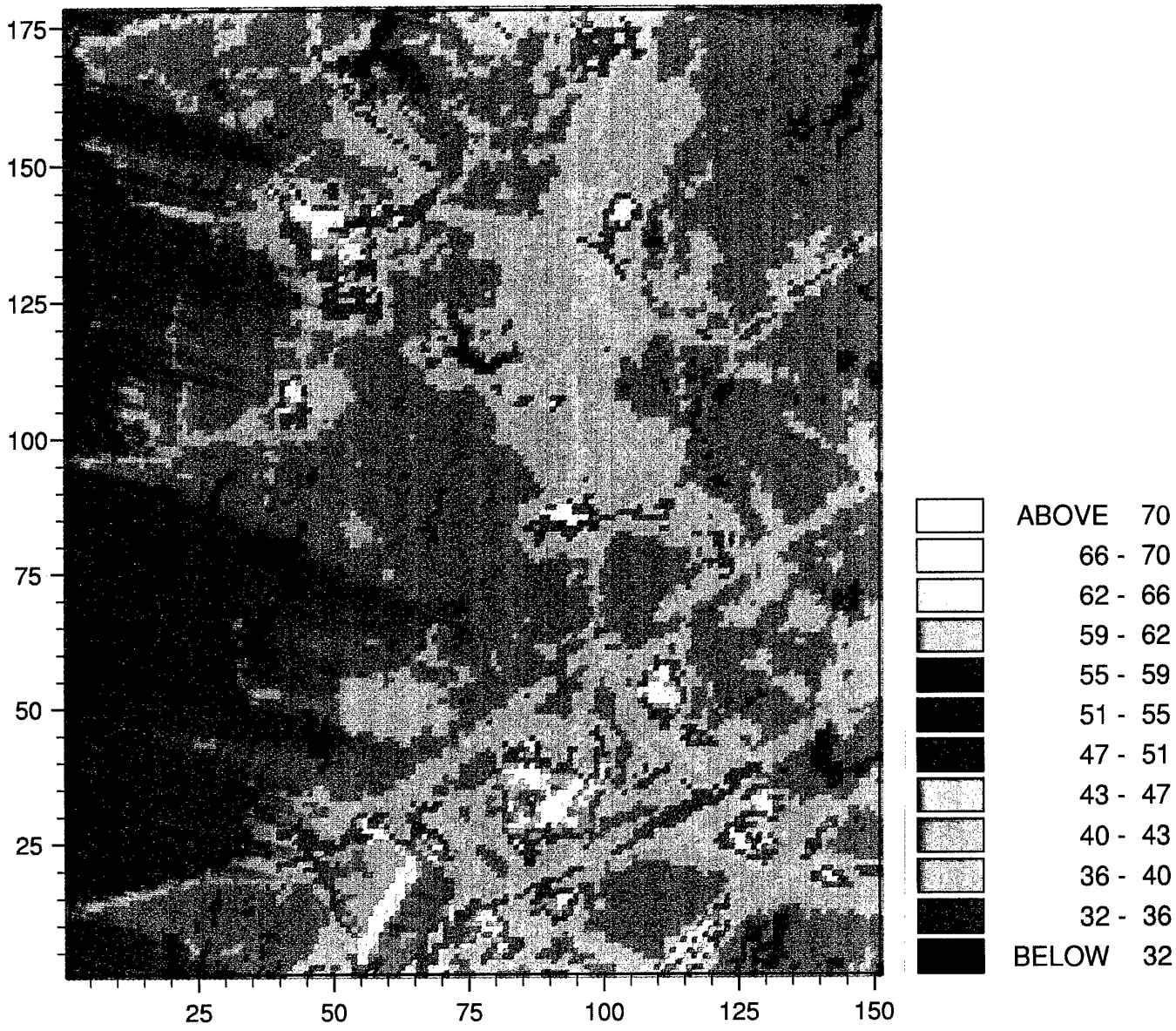

Figure 2: Pixel map of spectral values for channel 2 (Green) from SPOT data, A. P. Hill.

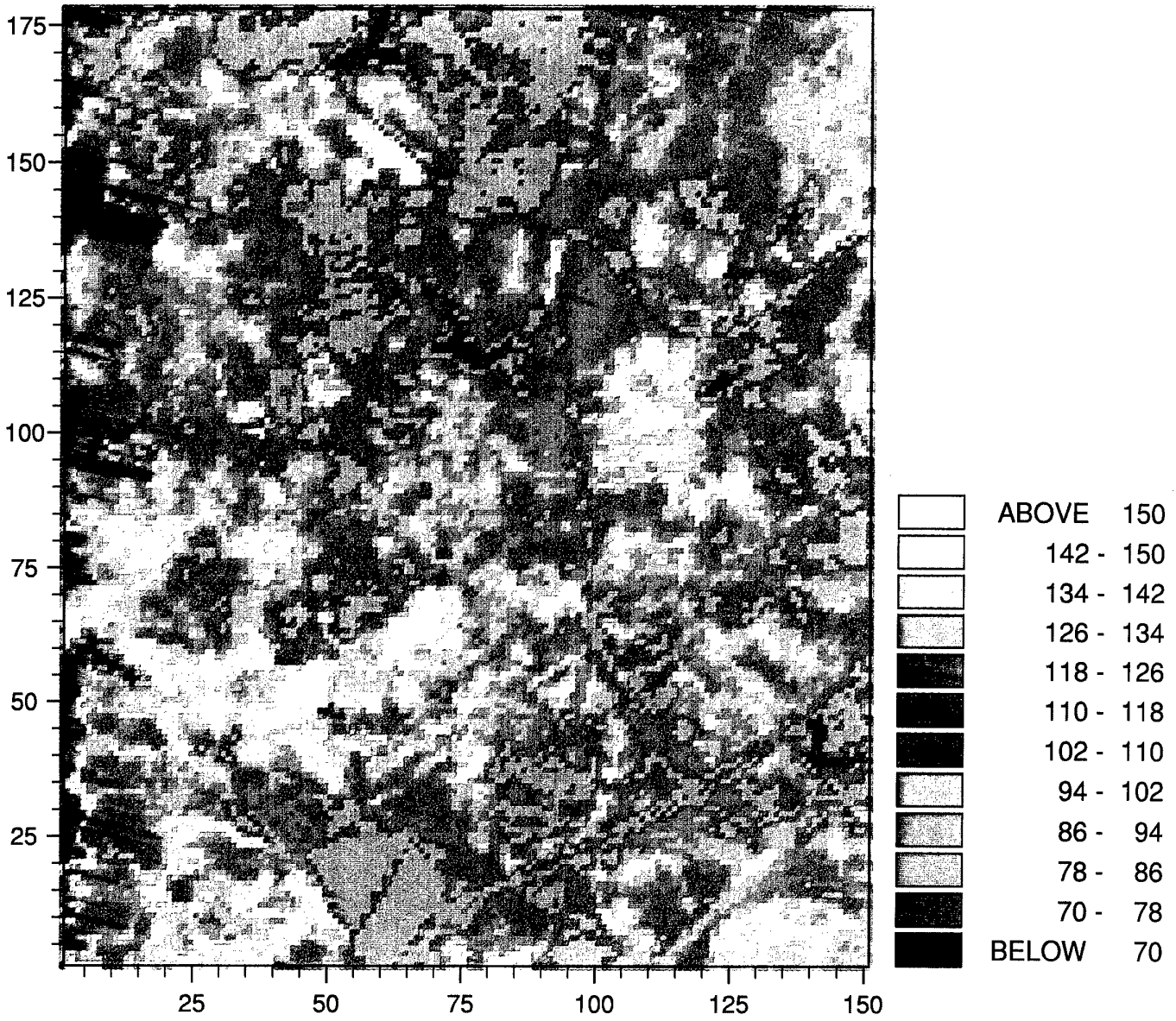

Figure 3: Pixel map of spectral values for channel 3 (NIR) from SPOT data, A. P. Hill.

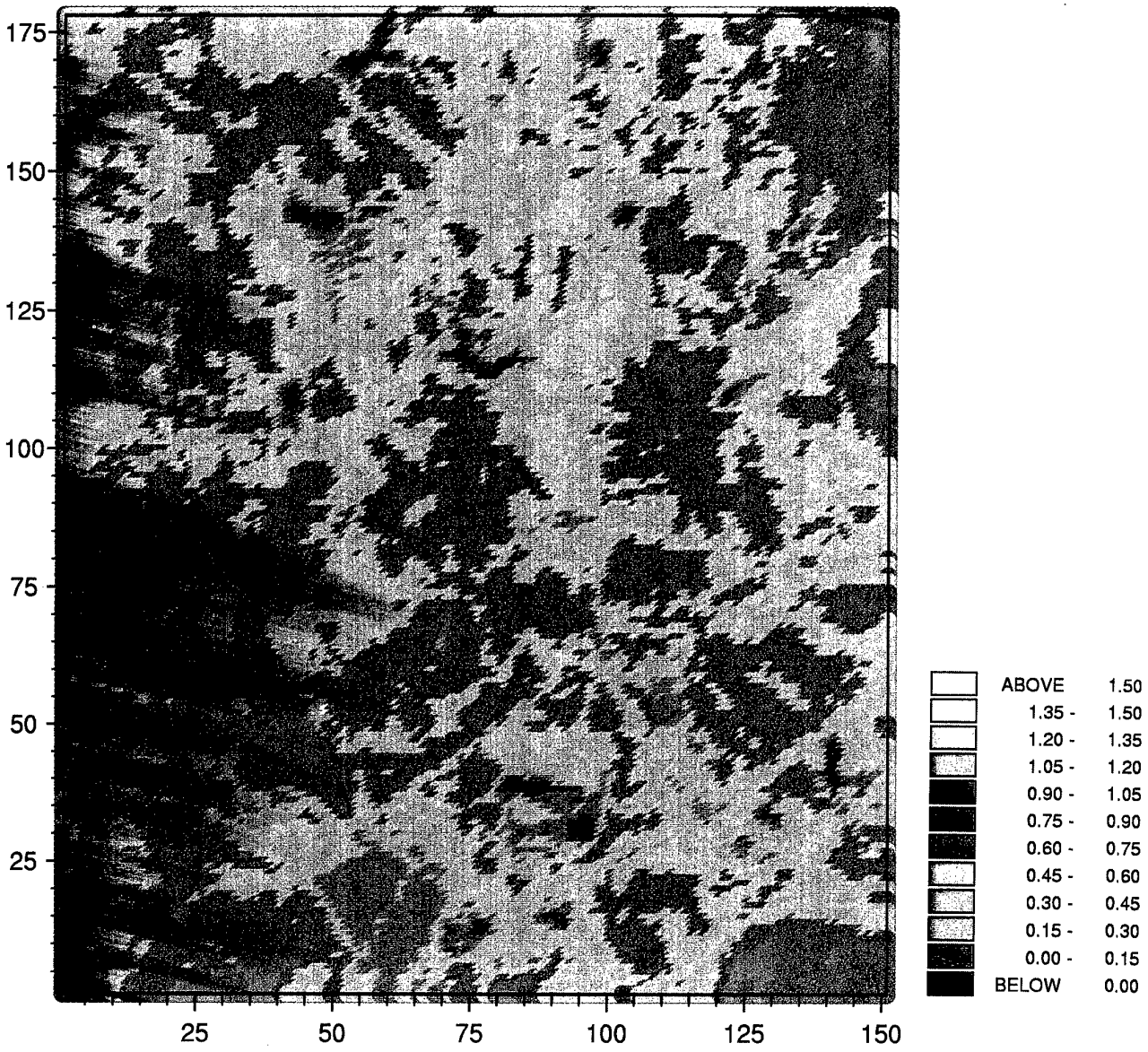

Figure 4: Pixel map of spectral values for NDVI from SPOT data, A. P. Hill.

PART II: VARIOGRAPHY, KRIGING AND KRIGING ANALYSIS OF SPOT IMAGERY

The theory and formulae for computing the variogram, for kriging and for filtering image data using kriging analysis were decribed in previous reports (PR-N68171-95-C-9007 and PR-N68171-95-C-9128). They are given in Appendix II of this report.

VARIOGRAPHY

Variograms were computed from the raw SPOT data for the rows and columns separately and also as averages of the rows and columns using the usual computing formula:

$$\widehat{\gamma}(\mathbf{h}) \quad = \quad \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2 \ .$$

where $\widehat{\gamma}(\mathbf{h})$ is the estimate of $\gamma(\mathbf{h})$ (the semivariance) at lag $\mathbf{h}$, $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i + \mathbf{h})$ are the observed values of Z in any one waveband at $\mathbf{x}_i$ and $\mathbf{x}_i + \mathbf{h}$, respectively, and $m(\mathbf{h})$ is the number of paired comparisons at that lag. By changing $\mathbf{h}$ we obtained a set of semivariances, which is the *experimental variogram* or *sample variogram.*

Mathematical models were then fitted to the experimental variograms using Genstat (Genstat 5 Committee, 1993). Several models were tried: circular, exponential, pentaspherical, power, spherical, and double (nested) exponential and spherical.

Variogram Results

Figures 5 and 6 show the experimental and model variograms for rows, columns, average and for each of the wavebands and for NDVI. Table 2 gives the best fitting models and their parameters. The experimental variograms for the three channels and NDVI for both rows and columns have a similar form—they all show strong evidence of spatial correlation, Figures 5 and 6. The best fitting model was principally the nested or double spherical model, but for some variograms of the southern part of the image it was the nested exponential one, Table 3 and Figure 7. The equation for the nested spherical model is:

$$\gamma(h) \quad = \quad c_0 + c_1 \left\{ \frac{3h}{2a_1} - \frac{1}{2} \left(\frac{h}{a_1} \right)^3 \right\} + c_2 \left\{ \frac{3h}{2a_2} - \frac{1}{2} \left(\frac{h}{a_2} \right)^3 \right\} \quad \text{for} \quad 0 < h \leq a_1 \quad (1)$$

$$\gamma(h) = c_0 + c_1 + c_2 \left\{ \frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3 \right\} \quad \text{for} \quad a_1 < h \leq a_2$$
$$\gamma(h) = c_0 + c_1 + c_2 \quad \text{for} \quad h > a_2,$$

where c_1 and a_1 are the sill and range of the short range component of the variation, and c_2 and a_2 are the sill and range of the long range component.

That for the nested exponential model is:

$$\gamma(h) = c_0 + c_1\{1 - \exp(-h/a_1)\} + c_2\{1 - \exp(h/a_2)\} .$$

For this model a_1 and a_2 are the distance parameters of the short- and long-range components, respectively. The distance parameters a_1 and a_2 may be multiplied by 3 to give approximate correlation ranges for these components. Of the several other models fitted none provided such a good fit as the two nested models.

All of the variograms for the full study area were fitted by a nested spherical model. They all show clearly that there are at least two ranges of spatial variation: a short-range one averaging 8 pixels (160 m) for the average variogram, and a long one of about 32 pixels (640 m), Table 2. The ranges for the NDVI are much the same as for the individual wave bands. The model parameters for the rows and columns are somewhat different, however, suggesting that there is some short-range anisotropy in the variation. For each waveband and NDVI the spatial variation has a longer range in the East to West direction than North to South. There is some slight evidence for this in the pixel maps of channels 1 and 2, but the variation does not appear to be distinctly anisotropic. The pixel map of channel 3 does not show any clear evidence of anisotropy, but that for NDVI shows slight evidence of directional variation.

Overall the results are consistent, which suggests that they are reflecting a real pattern in the ground cover. The graphs of the variograms for all of the above analyses look similar (Figures 5 to 7).

Variograms were also computed for the southern part of the study area and modelled. Table 3 gives the model parameters. Double spherical and double exponential models provided the best fit to different channels. The main short range components are 100 m and 400 m. A longer range component of about 1.75 km is now evident. On average the ranges for the southern part of the area are longer than those for the North.

Table 2: Variogram model parameters for the three channels and NDVI from SPOT data, A. P. Hill.

		Variance			Range/m	
Channel	Model	c_0	c_1	c_2	a_1	a_2
Red – rows	Double sph	1.883	15.682	14.127	318.90	812.12
– columns	Double sph	0.0	6.860	16.450	124.30	421.16
– average	Double sph	0.4661	8.080	18.432	165.56	518.94
Green – rows	Double sph	3.144	23.553	29.607	312.28	895.14
– columns	Double sph	0.0	28.432	11.057	122.92	472.42
– average	Double sph	2.062	16.568	28.437	224.32	650.48
NIR – rows	Double sph	10.090	175.18	104.57	151.74	993.94
– columns	Double sph	0.0	136.19	124.44	114.00	412.72
– average	Double sph	0.0	151.39	118.87	120.56	542.22
NDVI						
– rows	Double sph	0.00273	0.02404	0.02154	195.50	1077.76
– columns	Double sph	0.0	0.01716	0.02398	118.10	480.40
– average	Double sph	0.0	0.0343	0.01376	139.74	643.35

In the table c_0 is the nugget variance,
c_1 and c_2 are the sill variances of the short- and long-structures, respectively, and
a_1 and a_2 are the ranges of spatial dependence of the short- and long-structures. They are given in metres so that they can be compared with the results of the 1 m resolution imagery.

Table 3: Variogram model parameters for the three channels and NDVI from SPOT data for the southern part of the study area, A. P. Hill.

		Variance			Range/m	
Channel	Model	c_0	c_1	c_2	$a_1/3a_1$	$a_2/3a_2$
Red – average	Double sph	2.616	19.843	12.554	340.86	1438.22
Green – average	Double exp	0.0	33.696	35.246	438.78	2052.30
NIR – average	Double sph	0.0091	144.65	142.95	114.50	530.20
NDVI						
– rows	Double exp	0.0	0.03297	0.02186	131.68	1823.76
– columns	Double sph	0.0	0.02029	0.02739	136.66	634.96
– average	Double exp	0.000894	0.03789	0.01682	378.66	3011.04

In the table c_0 is the nugget variance,
c_1 and c_2 are the sill variances of the short- and long-structures, respectively, and
a_1 and a_2 are the ranges of the short- and long-structures for the spherical model or have been multiplied by 3 to obtain the effective ranges of spatial dependence for the exponential model.

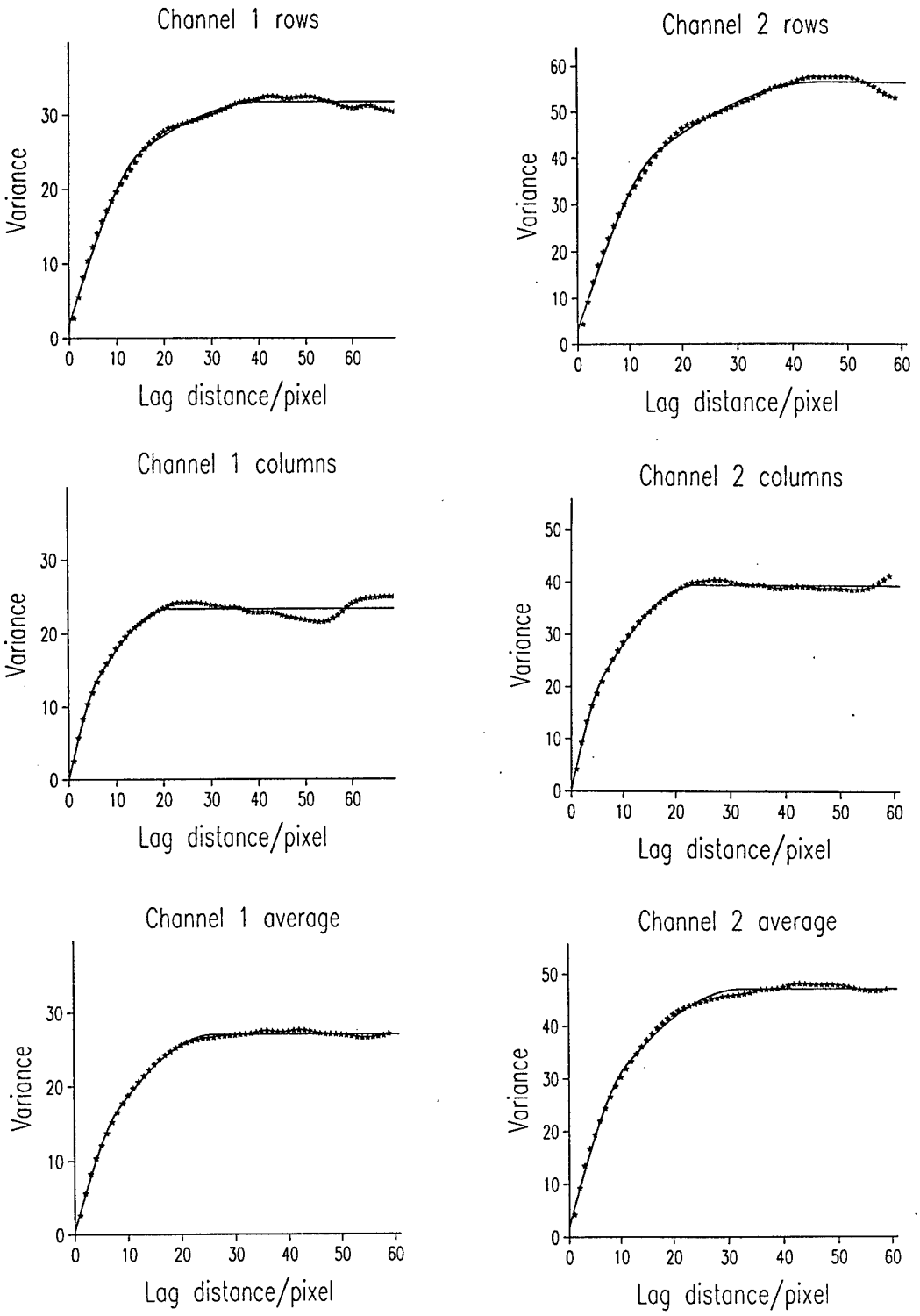

Figure 5: Variograms of rows, columns and the average for channels 1 (Red) and 2 (Green) from the SPOT data, A. P. Hill. The symbols are the experimental semi-variances, and the lines are the fitted models.

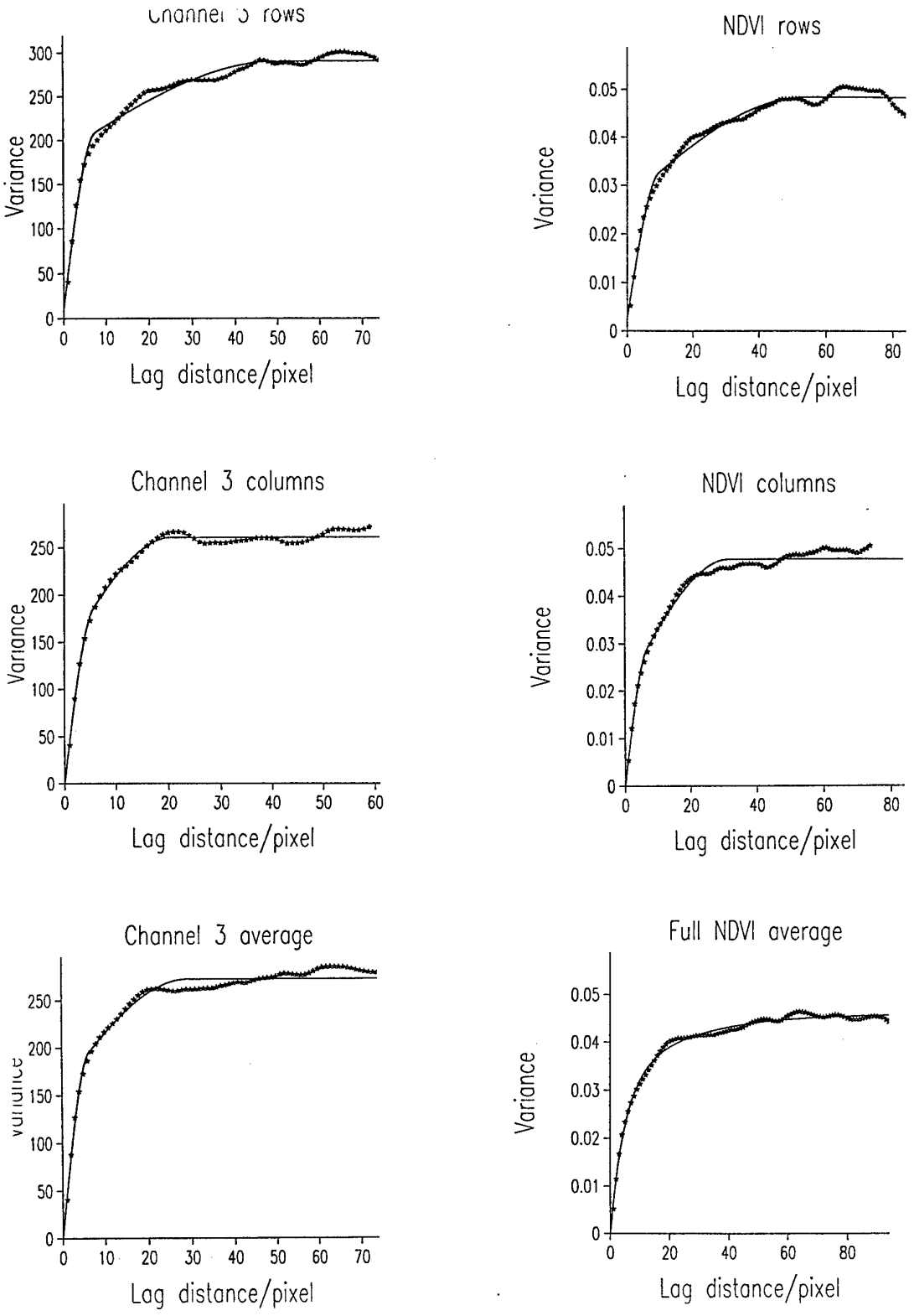

Figure 6: Variograms of rows, columns and the average for channel 3 (NIR), and NDVI from the SPOT data, A. P. Hill. The symbols are the experimental semivariances, and the lines are the fitted models.

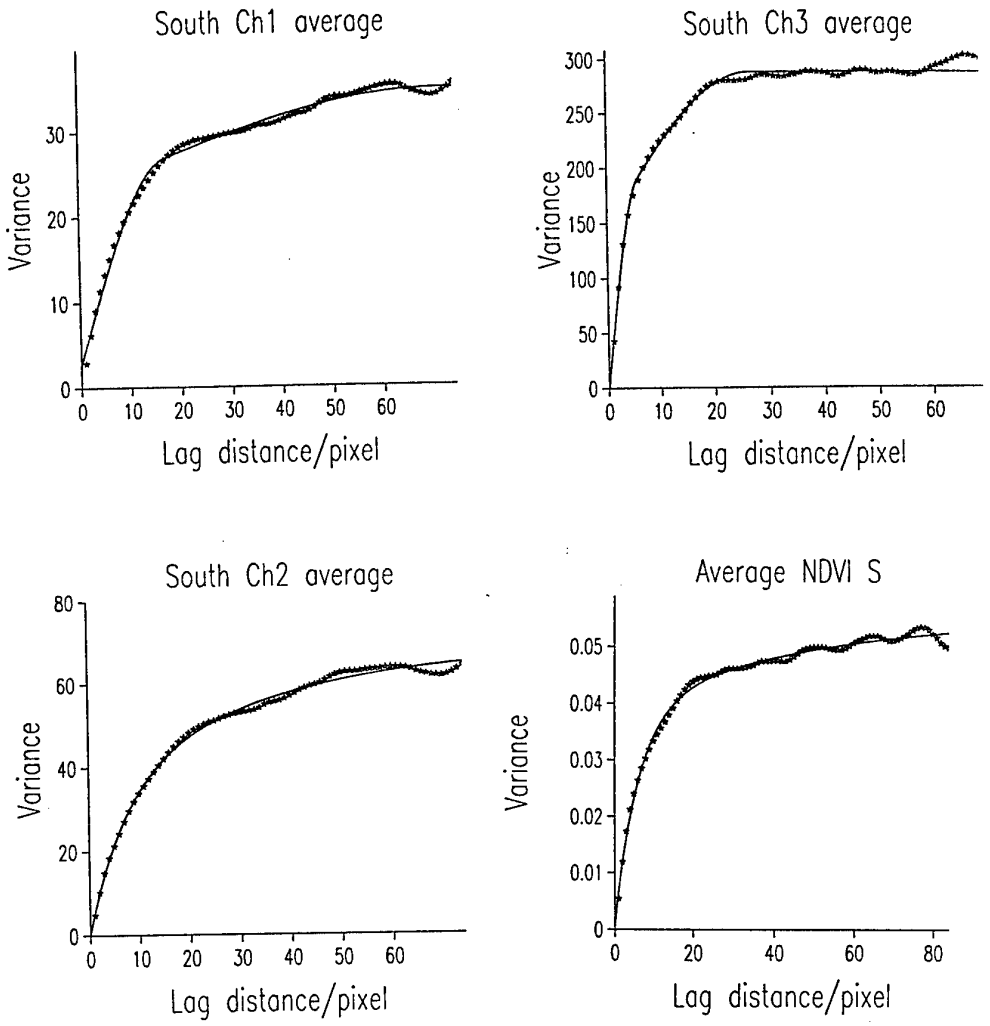

Figure 7: Average variograms for Channels 1, 2, and 3, and for NDVI from the SPOT data for the southern part of the study area, A. P. Hill. The symbols are the experimental semivariances, and the lines are the fitted models.

KRIGING AND KRIGING ANALYSIS

We used the parameters of the nested spherical model for the whole area for kriging and filtering the three wavebands and NDVI using the digitial numbers (DNs) of the SPOT image. The clear evidence that there are short- and long-range structures meant that filtering the scene was sensible to try to gain more insight into the variation present. We kriged the raw data on a grid that was offset by 0.5 of a pixel in both the north-south and east-west dimensions; this was to avoid estimating at the data points, which would have been pointless.

Results of ordinary kriging

Figures 8 to 11 show the kriged maps of the three channels and NDVI, respectively. The isarithmic intervals were chosen to reflect the variation in all bands to a similar extent. However, that for Channel 3 appears to be much greater.

The general appearance of the kriged maps for Channels 1 and 2 are remarkably similar, especially for the large reflectance values over the open spaces and roads (see southern part of maps around Anderson Camp). The map of Channel 3 is very different – much more variation is evident. The areas of high reflectance now have small DN values; in fact there is a reversal in the order of the reflectances. The kriged map of NDVI is very similar to that for Channel 3 with small values in the areas of high reflectance and large values over areas of woodland. It suggests that NDVI reflects the pattern in the Infra-red rather more than that in the red waveband. Neither of these maps show any evidence of anisotropy.

All of the kriged maps show clearly the short- and long-range components of the variation, especially those for NIR and NDVI. These are also the values that are the most stable statistically.

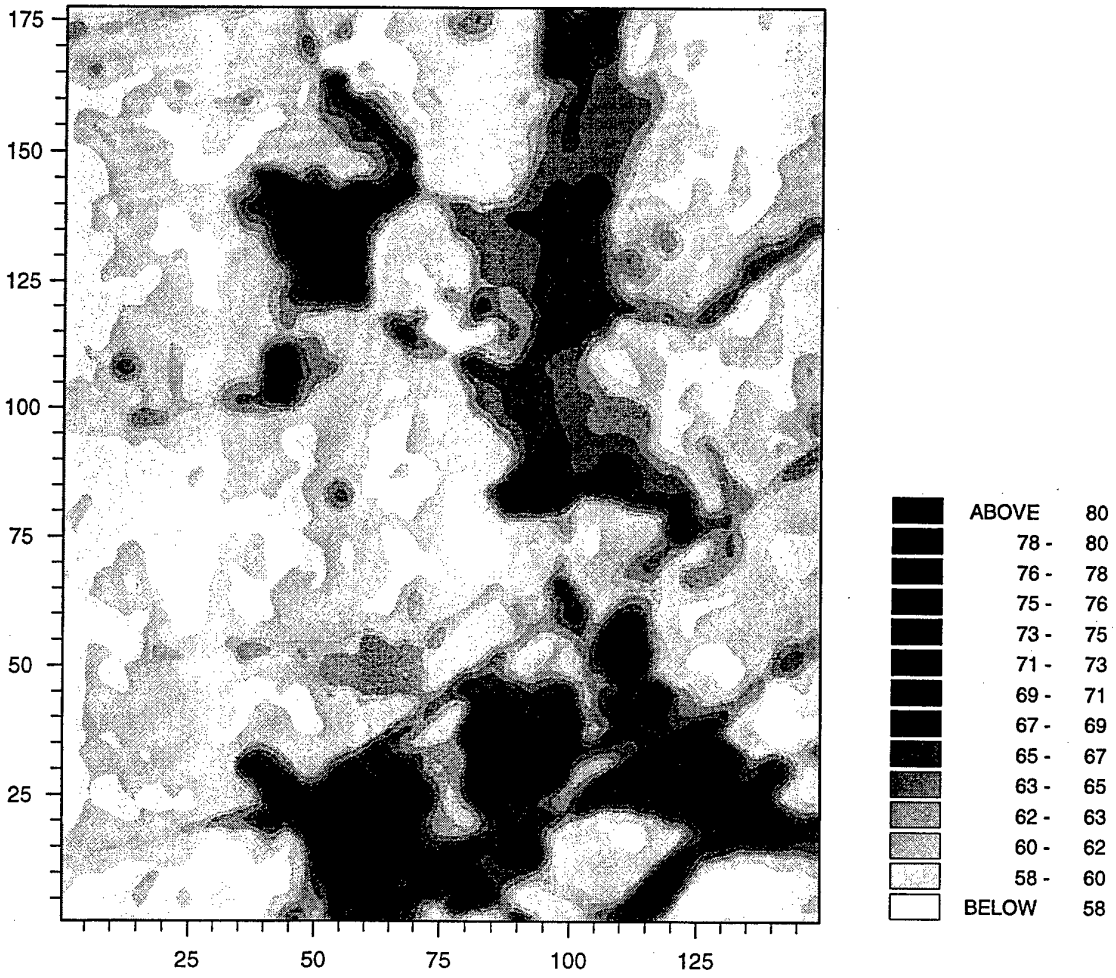

Figure 8: Map of ordinary kriged estimates for the A. P. Hill study area: Channel 1 (Red) SPOT data.

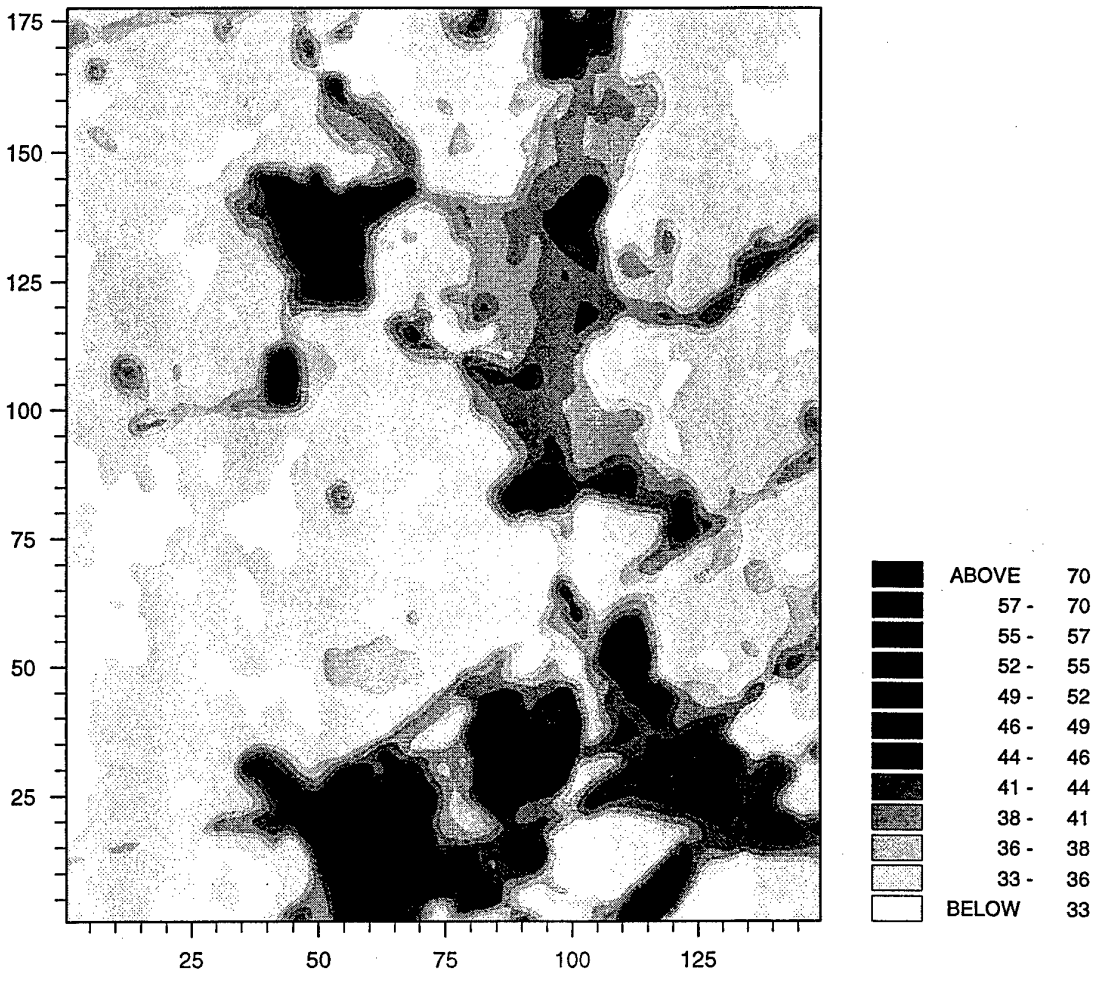

Figure 9: Map of ordinary kriged estimates for the A. P. Hill study area: Channel 2 (Green) SPOT data.

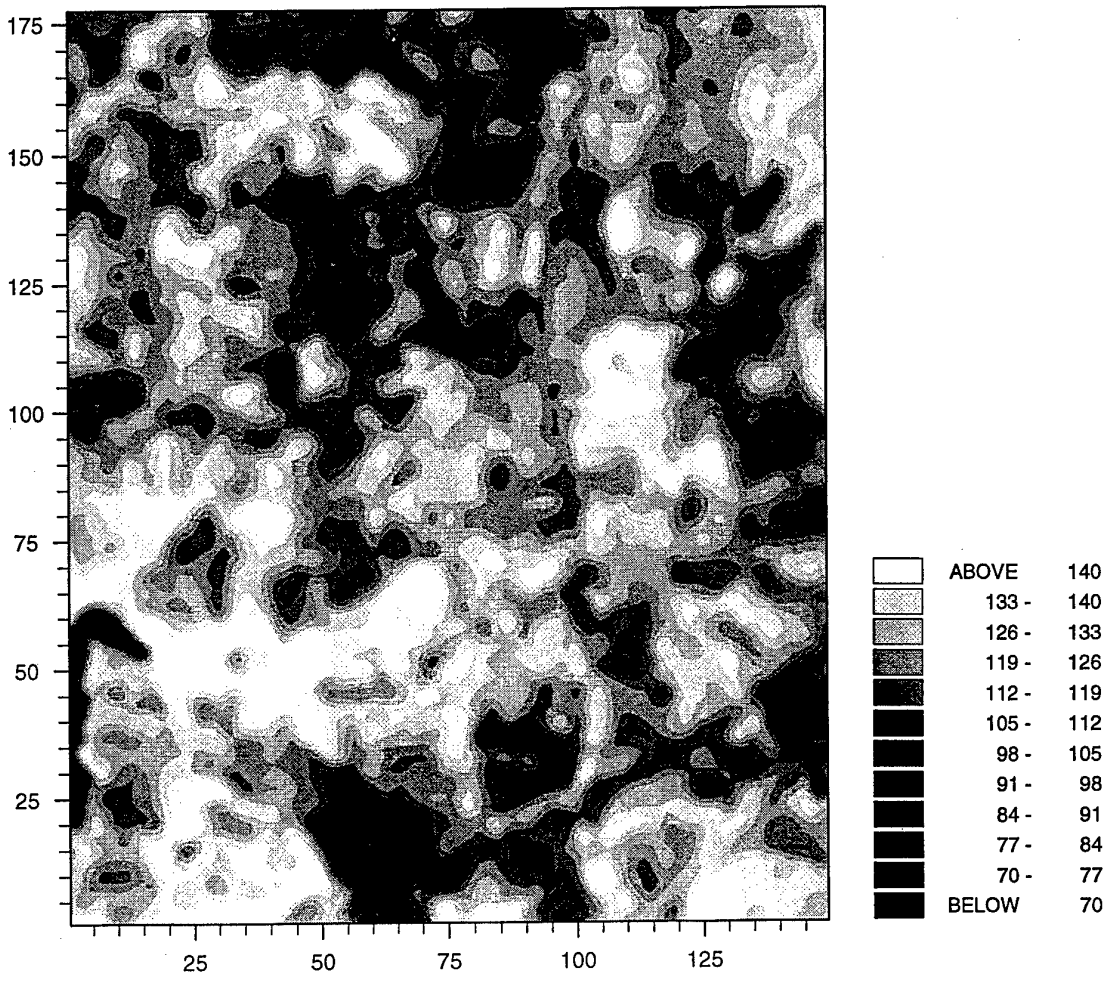

Figure 10: Map of ordinary kriged estimates for the A. P. Hill study area: Channel 3 (NIR) SPOT data.

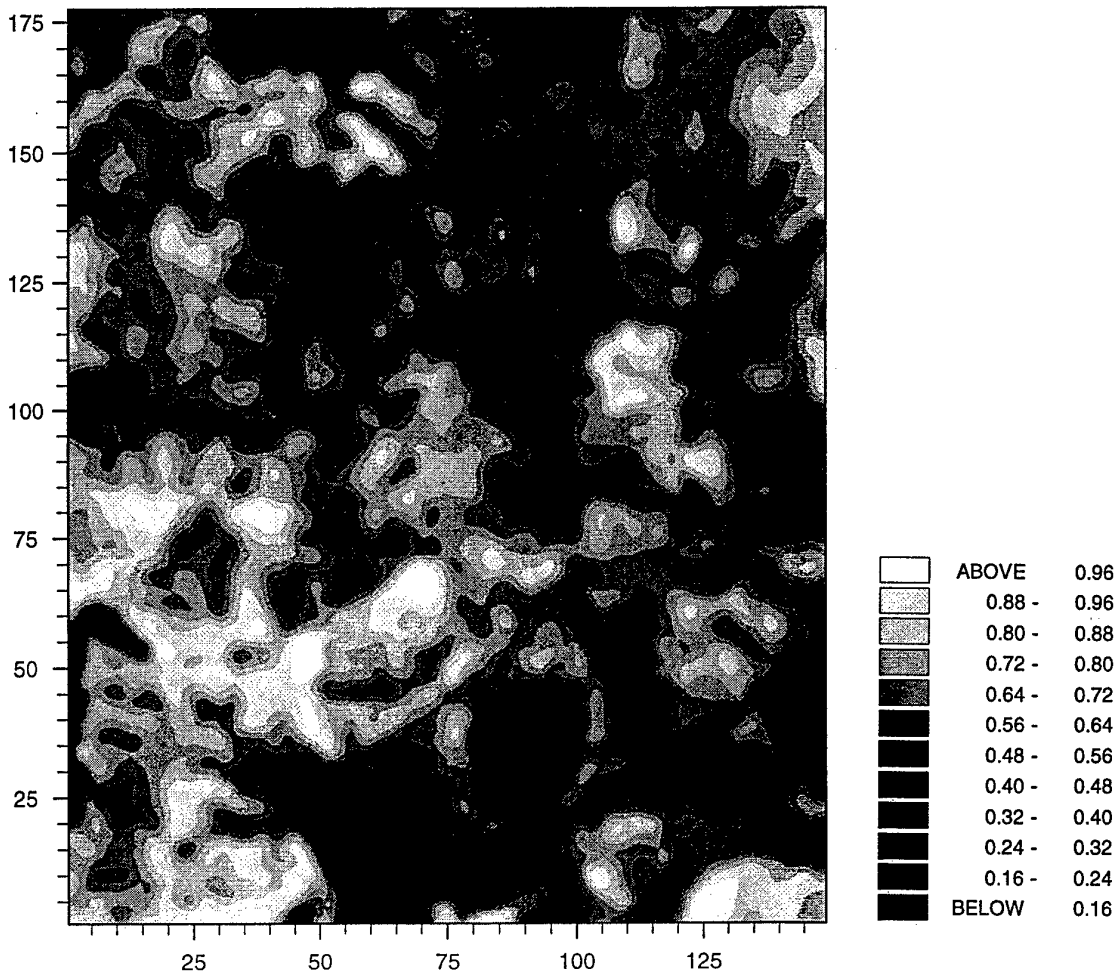

Figure 11: Map of ordinary kriged estimates for the A. P. Hill study area: NDVI, SPOT data.

Results of the kriging analysis: filtering

Long range

The long-range structures in the four maps, Figures 12 to 15, are clear. Kriging with the long-range filter evidently removes the local noise and probably shows the extent of the major ground cover classes. The maps for Channels 1 and 2 are very similar, they show the open areas, roads and areas of building clearly. These areas have the largest DNs. The map for channel 1 picks out clearly the valley bottoms and water bodies in dark blue (small DNs). The long-range map for Channel 3 shows more variation than those for the other channels. The areas of high reflectance and water bodies have small DNs whereas the vegetation in the valley bottoms has the largest DNs. The same kind of pattern is evident for the map of NDVI although it appears to be less variable, probably reflecting the longer range of the larger structure in the variogram. The large reflectance values are again in the valley bottoms, but the areas of open water have small values.

Short range

The maps of the short-range variation for Channels 1 and 2 (Figures 16 and 17) are very similar. Both show clearly the pattern of the roads, paths, buildings, and the airstrip. The cleared areas of grass are not particularly evident, but it is possible that some of these appear as areas of low reflectance near to those with high DNs. The maps of the short-range variation for Channel 3 (Figure 18) is perhaps the most interesting as it has identified the water bodies clearly as the areas of low reflectance values. They are more evident on this map than on that for NDVI (Figure 19). Both of these maps are similar, however, with the red areas in the valley bottoms and the green ones on the interfluves or spurs. The short-range maps for Channels 1 and 2, and those for Channel 3 and NDVI need to be examined together because it seems that the variation in the built-up and open areas is more evident in the former and the variation in the vegetation and drainage characteristics in the latter.

Summary

There is considerable detail evident on the maps of the long-range component, especially those for Channel 3 and NDVI. At present it is difficult to ascertain how they relate with those of the short range variation. The variation of the short-range component seems to relate to the detailed variation in physiography and drainage, whereas that of the long range seems to relate to the major land use types. Recent work by Wen and Sinding-Larsen (1997) has also shown the value of filtering imagery geostatistically.

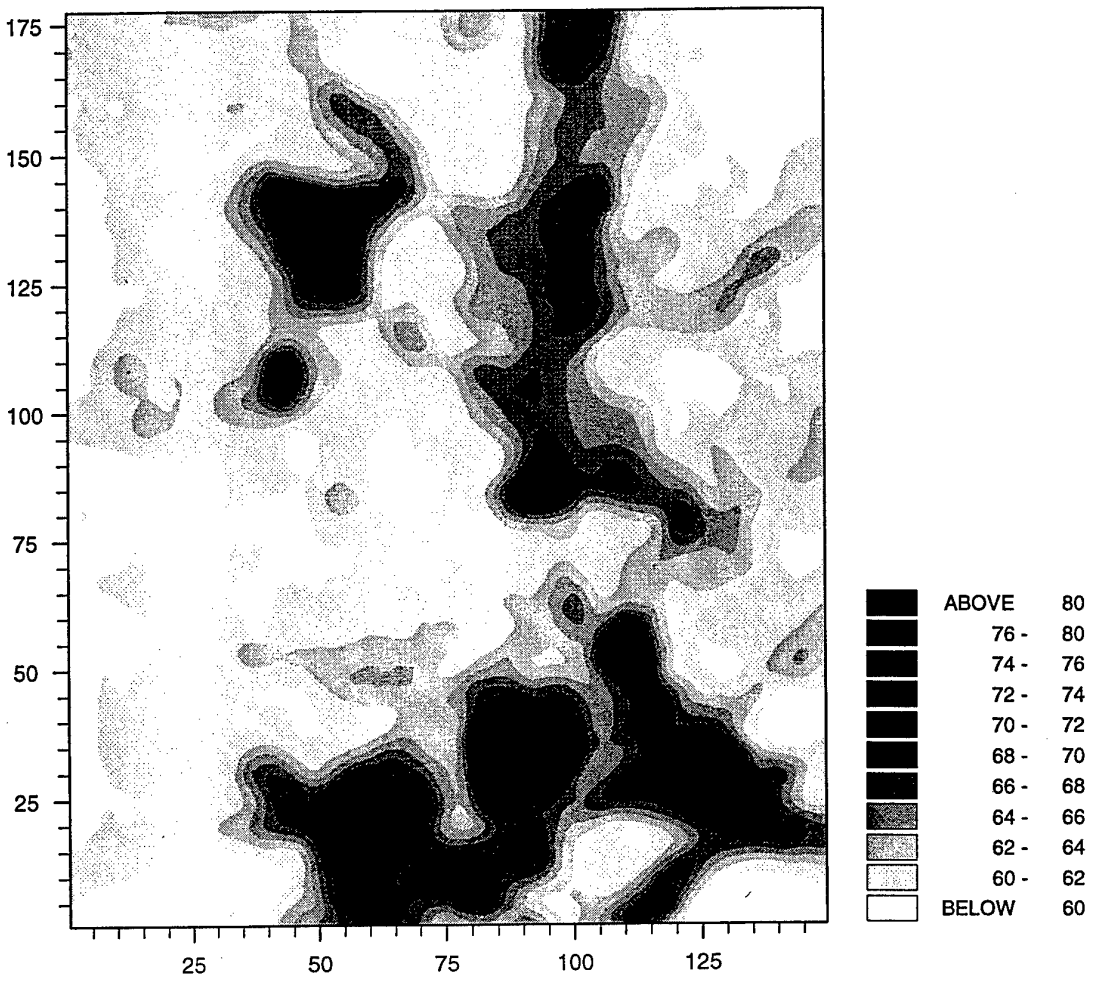

Figure 12: Map of long range kriged estimates after filtering for the A. P. Hill study area: Channel 1 (Red) SPOT data.

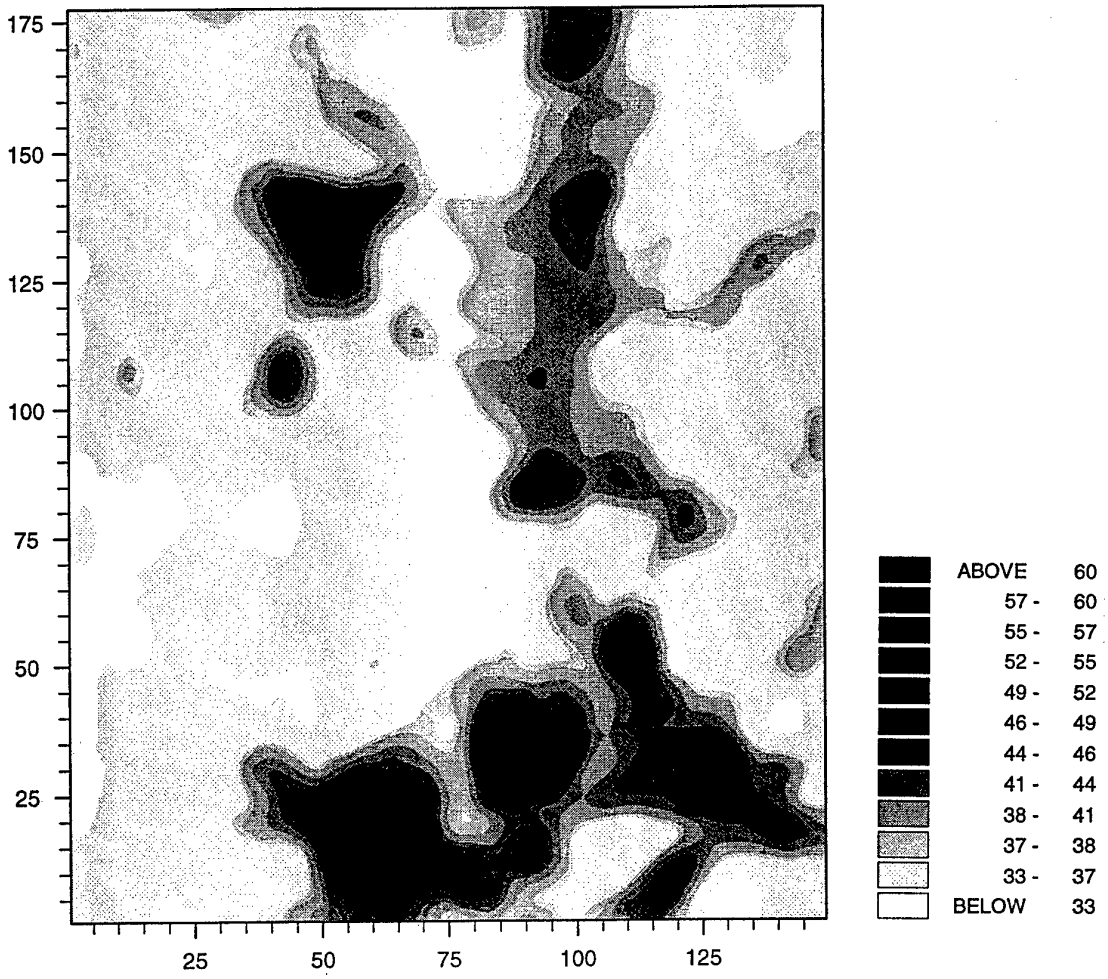

Figure 13: Map of long range kriged estimates after filtering for the A. P. Hill study area: Channel 2 (Green) SPOT data.

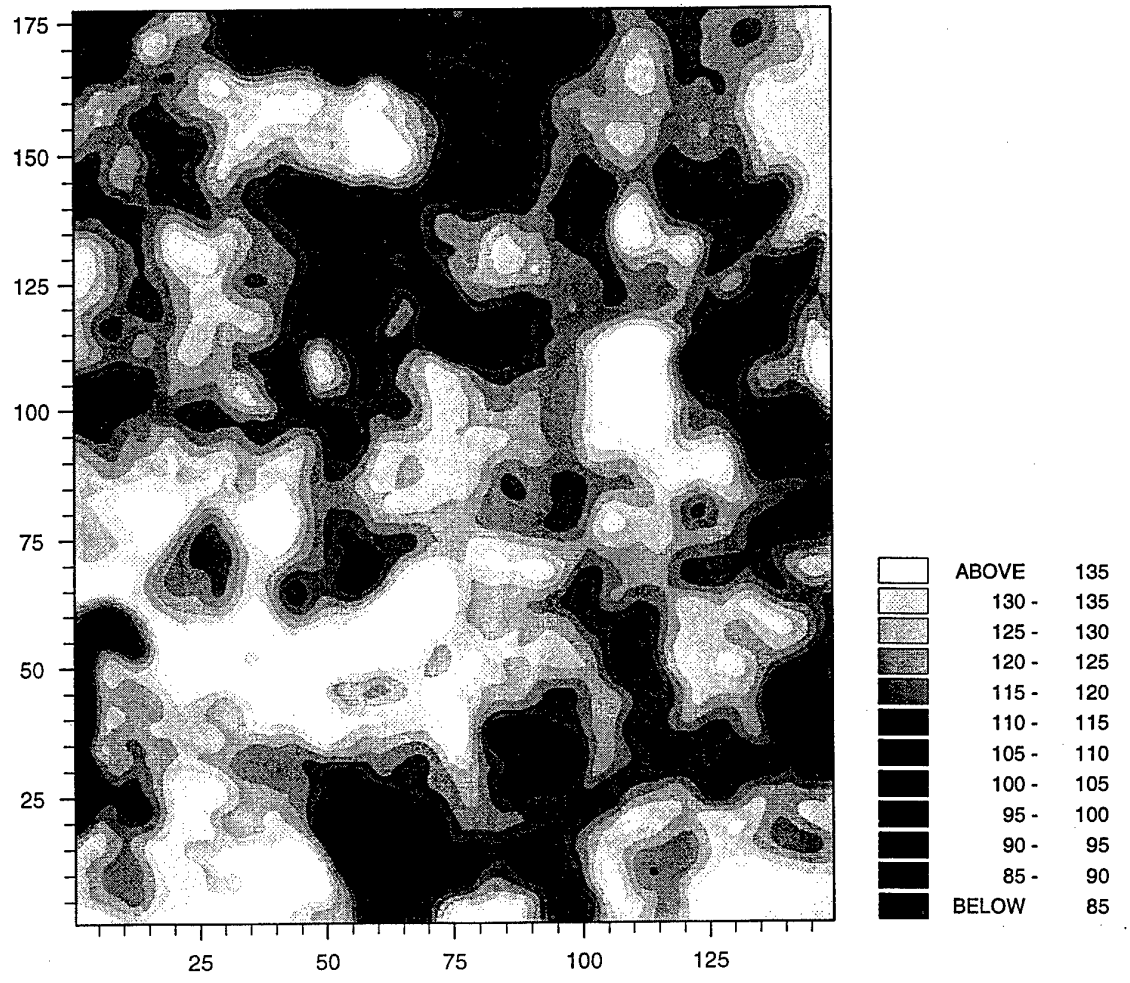

Figure 14: Map of long range kriged estimates after filtering for the A. P. Hill study area: Channel 3 (NIR) SPOT data.

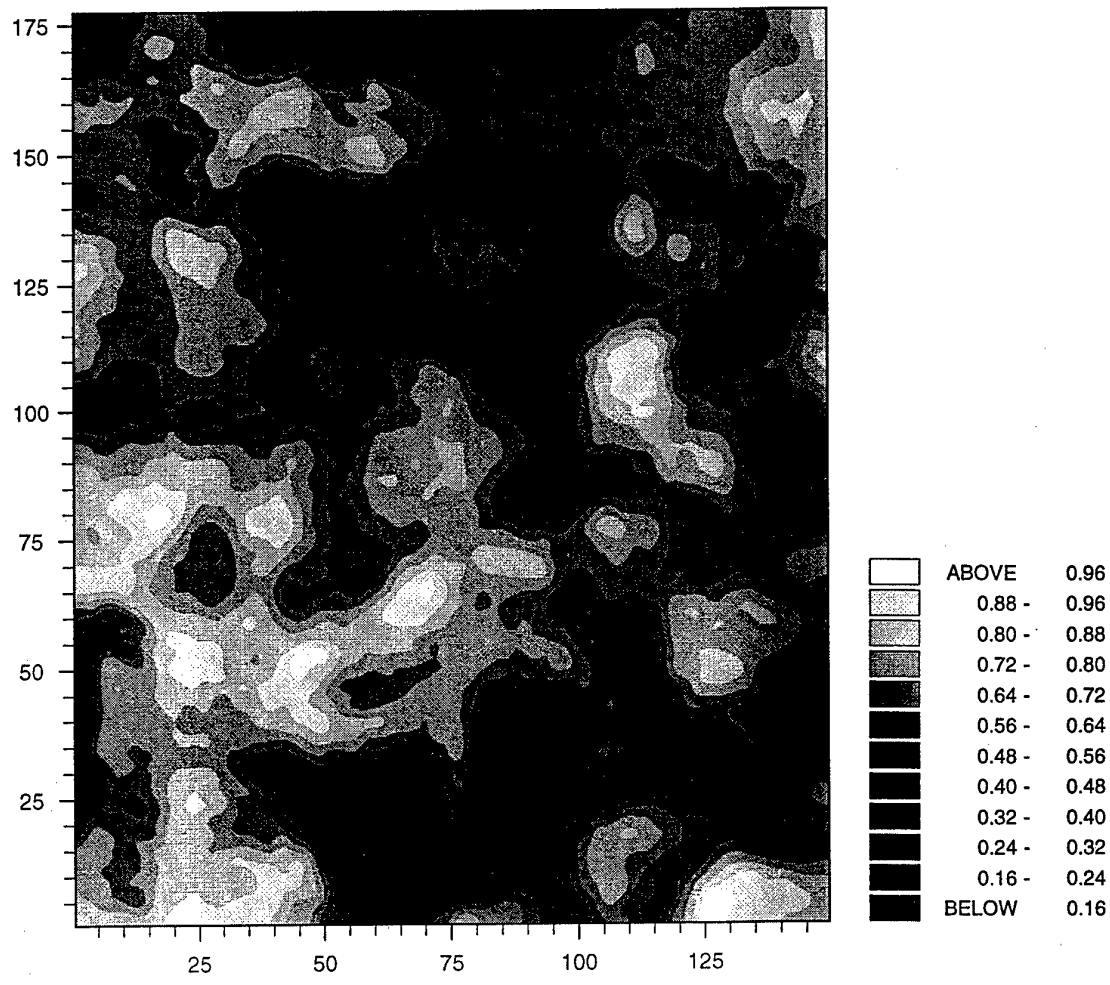

Figure 15: Map of long range kriged estimates after filtering for the A. P. Hill study area: NDVI, SPOT data.

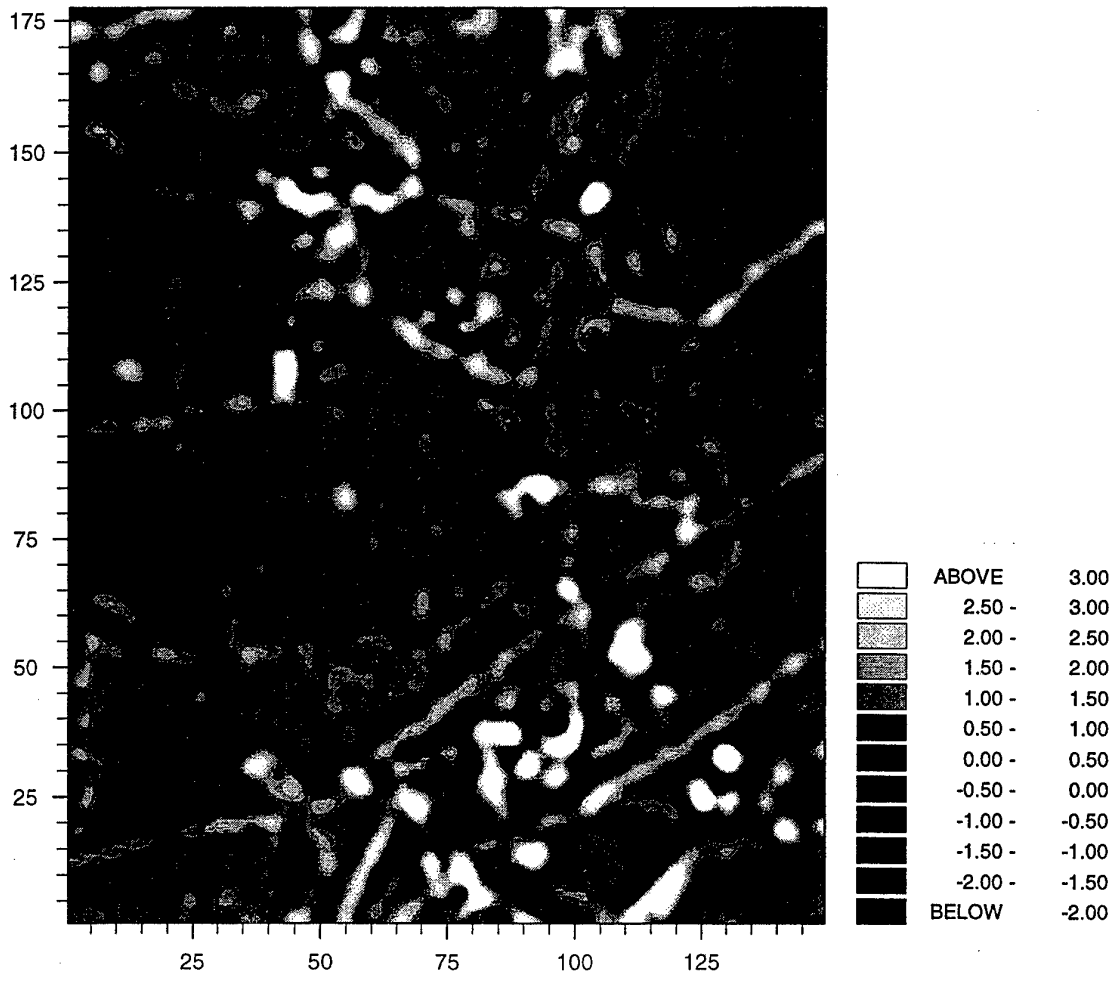

Figure 16: Map of short range kriged estimates after filtering for the A. P. Hill study area: Channel 1 (Red) SPOT data.

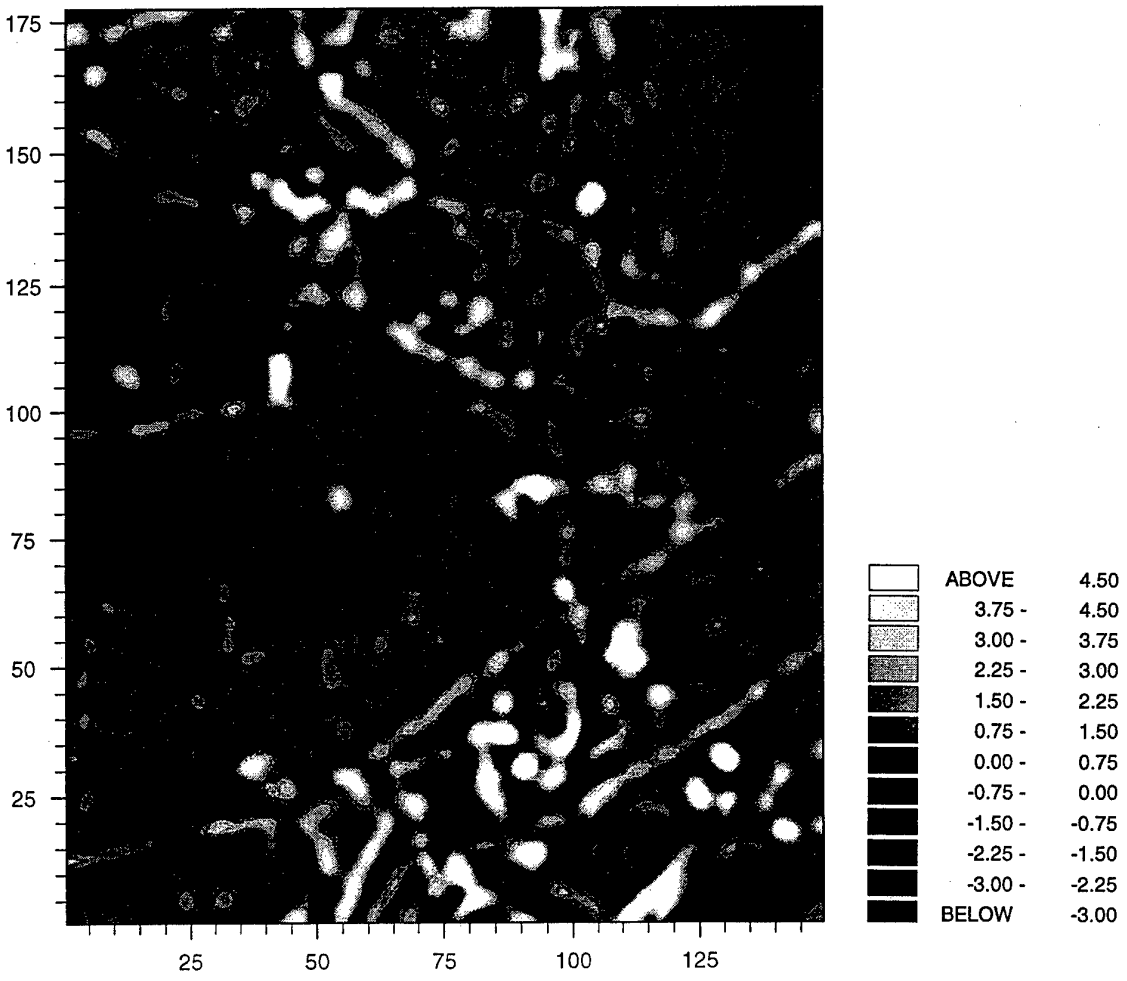

Figure 17: Map of short range kriged estimates after filtering for the A. P. Hill study area: Channel 2 (Green) SPOT data.

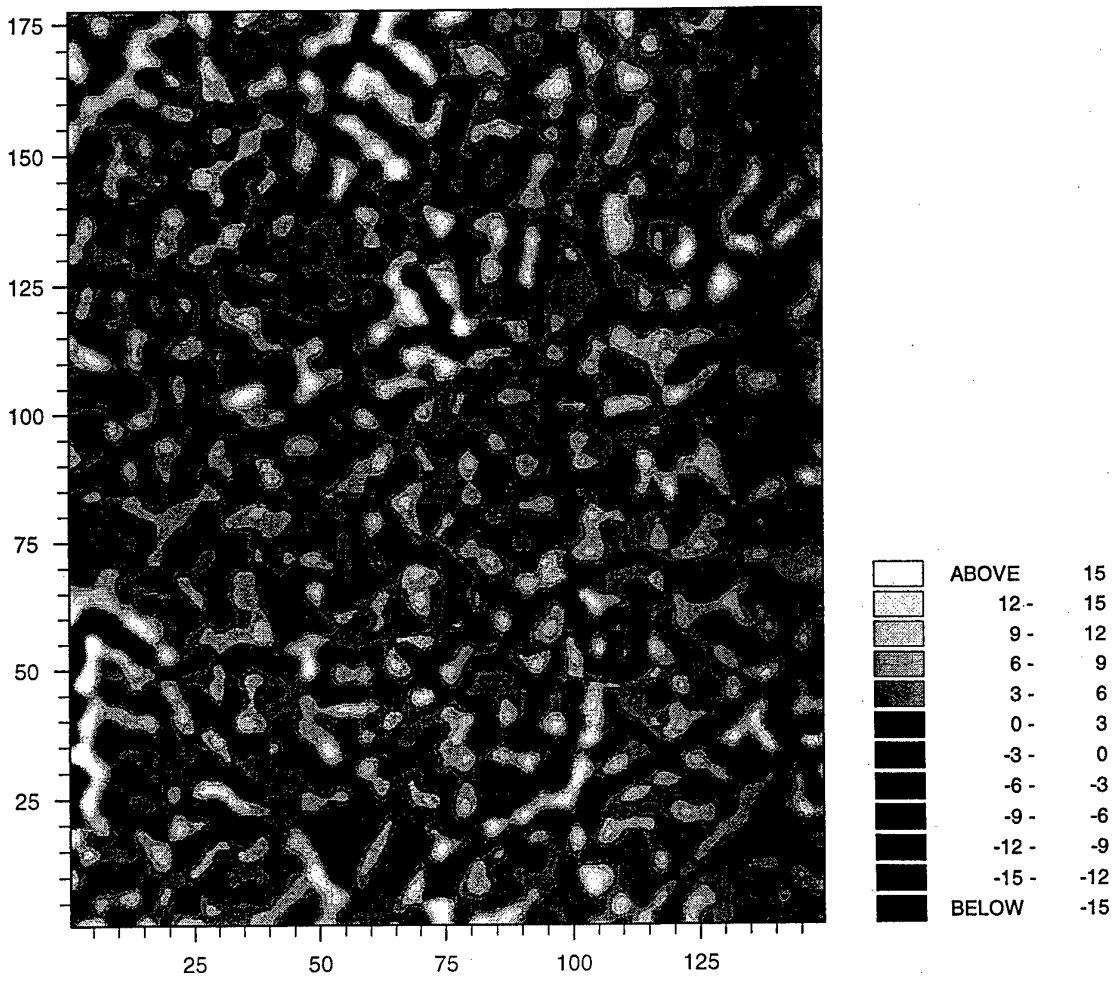

Figure 18: Map of short range kriged estimates after filtering for the A. P. Hill study area: Channel 3 (NIR) SPOT data.

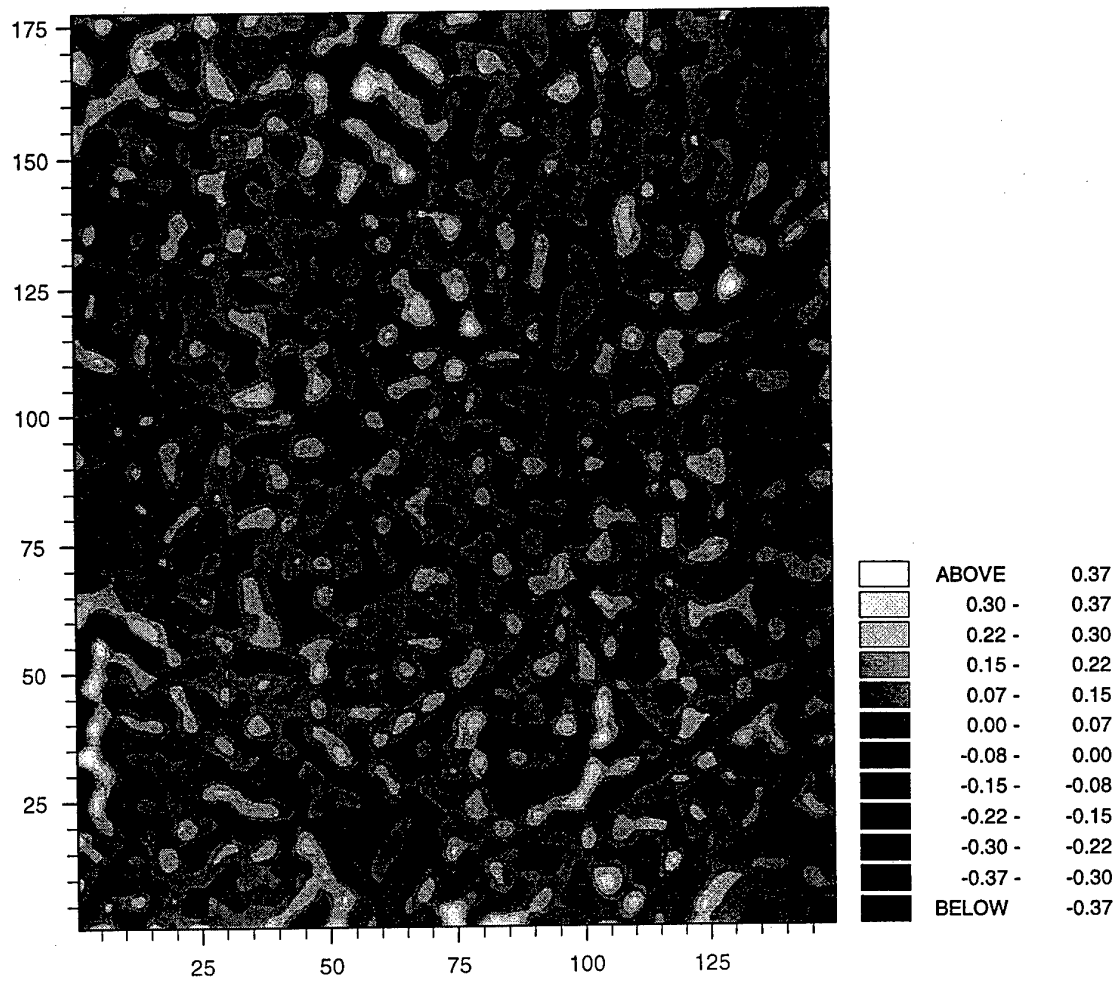

Figure 19: Map of short range kriged estimates after filtering for the A. P. Hill study area: NDVI, SPOT data.

PART III: GROUND COVER SURVEY

The results of the variogram and kriging analyses provided a basis for suggesting a sampling scheme to obtain information on the ground cover to relate to the information in the image. The ground information was obtained from quadrats along three transects that matched the SPOT pixels in both position and size, i.e. 20 m × 20 m. It was decided to record the cover of each pixel as a single class representing the dominant species. Since the short range component of the variation in the SPOT image was about 160 m, we suggested a sampling interval of 50 m which was well within this range. The long-range structure was about 640 m and the total transect length of 2500 m encompassed this fully. Hence this sampling scheme should encompass the main spatial patterns present on the ground.

Ground cover was recorded along two transects in the E-W direction and a third N-S. All had random starting positions. The ground cover classes included: hardwood, dense plantation, young plantation, shrubs, grass, mixed woodland, roads, softwoods, cleared areas and built up areas etc. In addition the pixel information for each of the ground survey points was also extracted from the image and analysed.

Ground Cover

The ground cover classes are listed in Table 4. There were several mixed pixels, and we had to judge to which class we should allocate them. For the first analysis we used 17 ground cover classes. This proved to be too many, because with only 165 sampling points several classes contained too few occurrences for a reasonable analysis. Therefore, in the second analysis we reduced the number of classes to seven. For the latter we also removed all mixed pixels from the data to see whether we could interpret the variation more clearly. These records were then transformed into presence/absence data for analysis.

Variogram Analysis and Results

Indicator variograms based on presence or absence scores were computed for each of the ground cover classes of the above groupings, i.e. 17 classes and 7 classes. These variograms are computed in the same way as those for the image data.

In addition the multivariate variogram was computed for both sets of classes. This requires a different method of analysis and the one that we used is that of Bourgault and Marcotte (1991). Its equation is

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}[\{\mathbf{Z}(\mathbf{x}) - \mathbf{Z}(\mathbf{x}+\mathbf{h})\}^{\mathrm{T}} \mathbf{M} \{\mathbf{Z}(\mathbf{x}) - \mathbf{Z}(\mathbf{x}+\mathbf{h})\}] , \qquad (2)$$

Table 4: Ground cover classes for the first three transect surveys, A. P. Hill.

Class	17 cover classes	Class	7 cover classes
1	Hardwood	1	Hardwood
2	Dense plantation	2	Dense plantation and softwood
3	Young plantation	3	Young plantation and shrubs
4	Softwood	4	Grass
5	Grass	5	Mixed vegetation
6	Shrubs	6	Roads, bivouac, industry
7	Mixed vegetation	7	Wetlands
8	Wetland with mixed		
9	Bivouac		
10	Gravel road		
11	Dirt road		
12	Wetland		
13	Open area		
14	Wetlands/hardwood		
15	Hardwood/field		
16	Sumac/grass		
17	Industry		

where $\mathbf{Z}(\mathbf{x})$ and $\mathbf{Z}(\mathbf{x}+\mathbf{h})$ are vectors of random variables at positions $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ separated by $\mathbf{h}$, the lag, and $\mathbf{M}$ is a $p \times p$ positive definite symmetric matrix defining the relations between the variables. For a regular one-dimensional sequence, such as our transect data, the experimental variogram can be calculated by the standard computing formula adapted for the multivariate case:

$$\widehat{\gamma}(h) \quad = \quad \frac{1}{2(n-h)} \sum_{i=1}^{n-h} \{\mathbf{z}(i) - \mathbf{z}(i+h)\}^{\mathrm{T}} \mathbf{M} \{\mathbf{z}(i) - \mathbf{z}(i+h)\} \ , \qquad (3)$$

in which the lag is now a scalar, $h = |\mathbf{h}|$, $\widehat{\gamma}(h)$ is the experimental semivariance at that lag, and $\mathbf{z}(i)$ and $\mathbf{z}(i+h)$ are the vectors of observations in the ith and $(i+1)$th segments respectively.

Figure 20 shows the experimental and model variograms for those ground cover classes for which they were reasonable. For all of the other classes the variograms seemed to be pure nugget because of the small number of occurrences or because of the extent of fragmentation of the classes along the transects. With a larger sample more of the ground cover classes might have shown spatial dependence. Figure 21a shows

the experimental multivariate variogram computed from 17 ground cover classes, and the fitted model. It is rather erratic possibly because of the large number of classes. Table 5 gives the best fitting models and their parameters. The ground cover classes showing the greatest degree of continuity were hardwood, young plantation and grass.

The best fitting models were as follows.

The nested (or double) spherical:

$$\begin{aligned}
\gamma(h) &= c_0 + c_1\left\{\frac{3h}{2a_1} - \frac{1}{2}\left(\frac{h}{a_1}\right)^3\right\} + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \text{ for } 0 < h \le a_1 \\
&= c_0 + c_1 + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \text{ for } a_1 < h \le a_2 \\
&= c_0 + c_1 + c_2 \text{ for } h > a_2 \\
&= 0 \text{ for } h = 0\,, \qquad (4)
\end{aligned}$$

where c_1 and a_1 are the sill and range of the short-range component of the variation, and c_2 and a_2 are the sill and range of the long-range component.

A periodic model with a power component:

$$\begin{aligned}
\gamma(h) &= c_0 + A\cos(2\pi h/\lambda) + B\sin(2\pi h/\lambda) + mh^\alpha \text{ for } 0 < h \\
&= 0 \text{ for } h = 0\,, \qquad (5)
\end{aligned}$$

in which the parameters m and α are the slope and the exponent of the power function, respectively, A, B determine the amplitude and phase, and λ is the length of the wave.

The circular model:

$$\begin{aligned}
\gamma(h) &= c_0 + c_1\left\{1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a_1}\right) + \frac{2h}{\pi a_1}\sqrt{1 - \frac{h^2}{a_1^2}}\right\} \text{ for } 0 < h \le a_1 \\
&= c_1 \text{ for } h > a_1 \\
&= 0 \text{ for } h = 0\,, \qquad (6)
\end{aligned}$$

in which the parameters c_1 and a_1 are again the sill and range.

We have added a nugget variance, c_0, to all three.

the experimental multivariate variogram computed from 17 ground cover classes, and the fitted model. It is rather erratic possibly because of the large number of classes. Table 5 gives the best fitting models and their parameters. The ground cover classes showing the greatest degree of continuity were hardwood, young plantation and grass.

The best fitting models were as follows.

The nested (or double) spherical:

$$\begin{aligned}
\gamma(h) &= c_0 + c_1\left\{\frac{3h}{2a_1} - \frac{1}{2}\left(\frac{h}{a_1}\right)^3\right\} + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \text{ for } 0 < h \le a_1 \\
&= c_0 + c_1 + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \text{ for } a_1 < h \le a_2 \\
&= c_0 + c_1 + c_2 \text{ for } h > a_2 \\
&= 0 \text{ for } h = 0\,, \qquad (4)
\end{aligned}$$

where c_1 and a_1 are the sill and range of the short-range component of the variation, and c_2 and a_2 are the sill and range of the long-range component.

A periodic model with a power component:

$$\begin{aligned}
\gamma(h) &= c_0 + A\cos(2\pi h/\lambda) + B\sin(2\pi h/\lambda) + mh^\alpha \quad \text{for} \quad 0 < h \\
&= 0 \quad \text{for} \quad h = 0\,, \qquad (5)
\end{aligned}$$

in which the parameters m and α are the slope and the exponent of the power function, respectively, A, B determine the amplitude and phase, and λ is the length of the wave.

The circular model:

$$\begin{aligned}
\gamma(h) &= c_0 + c_1\left\{1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a_1}\right) + \frac{2h}{\pi a_1}\sqrt{1 - \frac{h^2}{a_1^2}}\right\} \quad \text{for} \quad 0 < h \le a_1 \\
&= c_1 \quad \text{for} \quad h > a_1 \\
&= 0 \quad \text{for} \quad h = 0\,, \qquad (6)
\end{aligned}$$

in which the parameters c_1 and a_1 are again the sill and range.

We have added a nugget variance, c_0, to all three.

Table 5: Variogram model parameters for the ground cover from transect surveys at A. P. Hill with 17 classes.

		Parameters				
Ground cover	Model	c_0	c_1	c_2	a_1/m	a_2/m
Multivariate						
All classes	Double sph	4.6366	1.3729	0.5724	202.0	723.0
Young plantation	Double sph	0.0111	0.0278	0.0174	295.5	1040.0
Hardwood	Circular	0.1181	0.0525		435.0	
Softwood	Pure nugget	Variance				

		Parameters					
Ground cover	Model	c_0	m	A	B	α	λ/m
Dense plantation	Periodic with power	0.0946	0.000062	−0.0077	0.0019	1.9161	400.0
Grass	Periodic with power	0	0.0784	−0.0222	−0.0164	0.1875	760.0

In the table c_0 is the nugget variance,
c_1 and c_2 are the sill variances of the short- and long-structures, respectively, and
a_1 and a_2 are the distance parameters of the short- and long-range structures.
For the circular model there is only one structure with parameters c_1 and a_1.
For the periodic with power function m is the gradient and α is the exponent of the power component, and A and B determine the amplitude and phase and λ is the wavelength (the average distance at which repetition in the ground cover occurs).

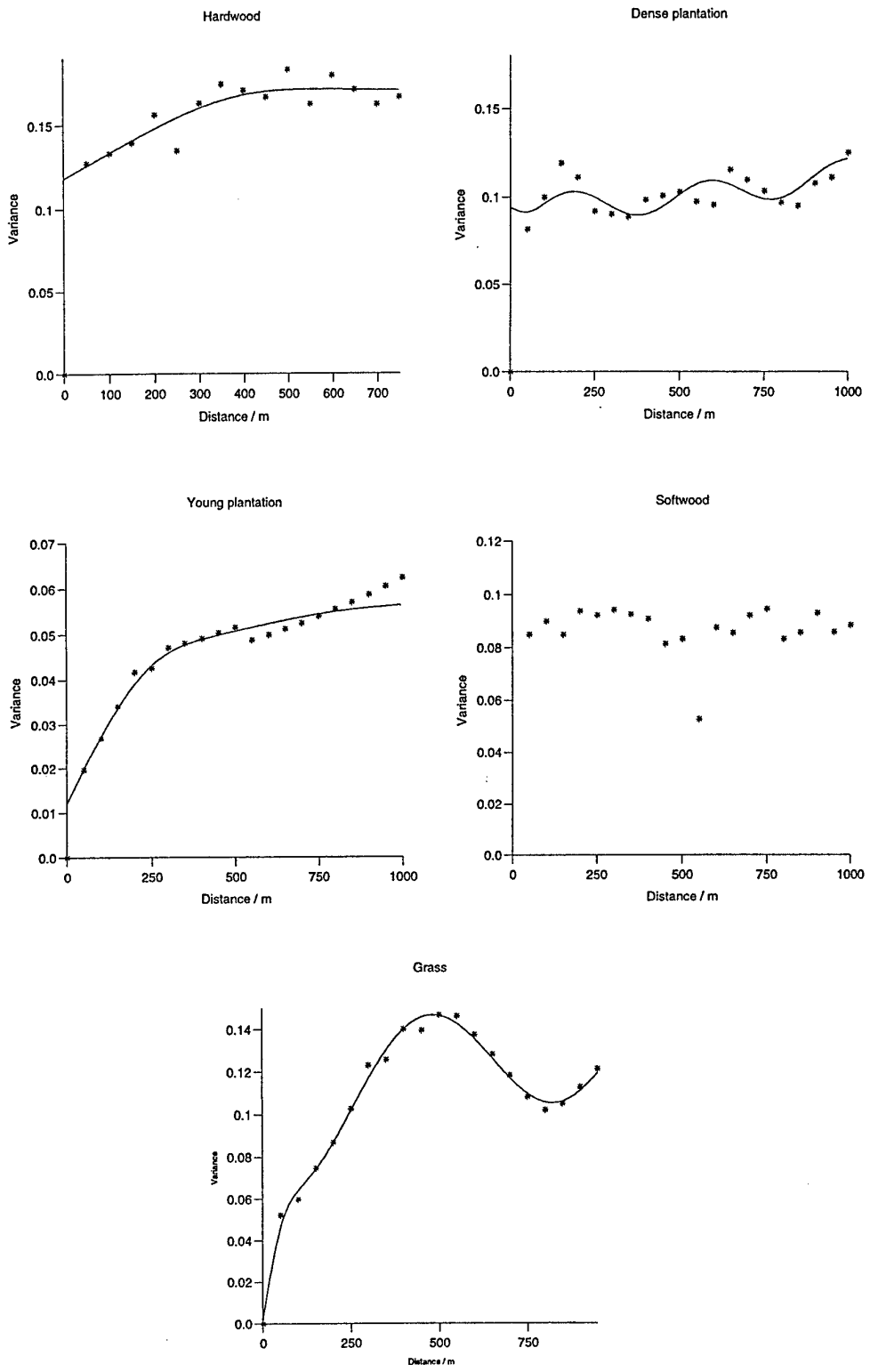

Figure 20: Variograms from transect data of ground cover for 17 classes at A. P. Hill. The symbols are the experimental semivariances, and the lines are the fitted models.

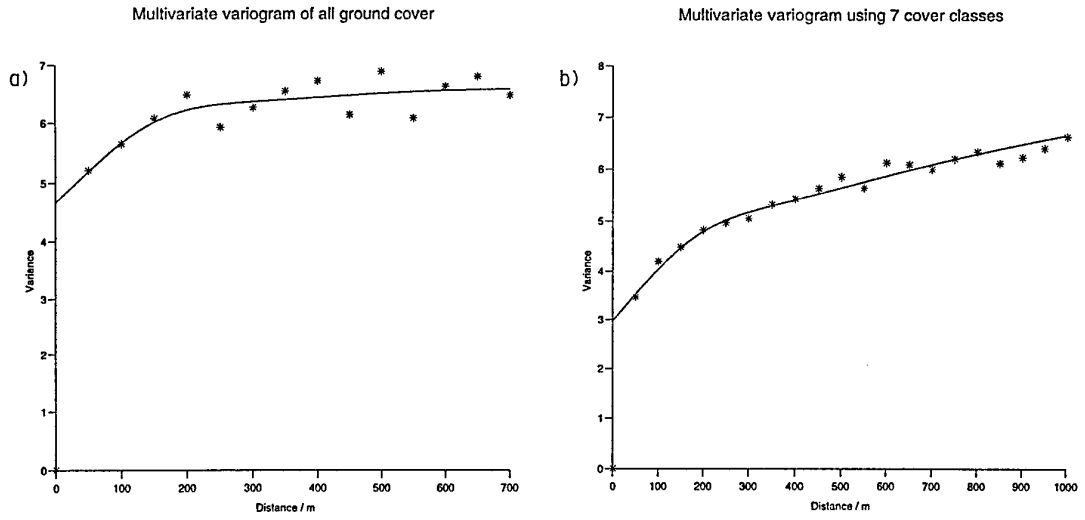

Figure 21: Multivariate variogram from transect data of ground cover at A. P. Hill: a) for 17 classes, and b) for 7 classes. The symbols are the experimental semivariances, and the lines are the fitted models.

The average scales of variation that appear to be significant are about 400 m for hardwood and dense plantation, and 700 m for grass and the long-range component of the multivariate variogram. The multivariate variogram (17) was fitted by a double spherical model with a short-range component of 200 m and a long-range one of about 725 m. This is close to the average range of the variograms from the full pixel information (Table 2).

When the number of classes was reduced to seven and all mixed classes were removed the results were similar to those for the main classes in the previous analysis. Figure 22 shows the variograms of the classes that were modelled and Table 6 gives the models and the parameters. The multivariate variogram computed with 7 classes is shown in Figure 21b, and its model parameters are given in Table 6. With seven ground cover classes a range of about 500 to 700 m appears to be important; the multivariate variogram, however, suggests ranges of 200 m and a much longer one of about 1500 m.

We extracted the pixel information from the SPOT image that coincided with the ground cover records for the N-S transect, and the E-W transect that was in our study area. Variograms were computed and modelled for each of the three wavebands and NDVI. The variograms are more erratic than those computed from the full set of data; it probably reflects the effect of the smaller number of data values and therefore less reliable estimates of the semivariances. Figure 23 shows the variograms and Table 7 gives the model parameters. The overall ranges of spatial dependence identified in the ground cover are reinforced by the variogram analysis of the pixel information from the transects. The 500 m scale of variation is evident from fitting the periodic with power function. We have also fitted simpler bounded models, and a longer range of variation of about 1200 m is evident. The circular model fitted to NIR has a range of about 200 m. With the additional ground cover information that TEC has provided we should be able to assess whether the three ranges of about 200 m, 500 to 700 m and 1200 to 1500 m are reflected in other parts of the region in the next project. The 200- m range corresponds with the short-range component of the multivariate variogram and of the overall pixel analysis. The 500 to 700-m range is evident from the full pixel analysis and from some of the ground cover classes. The longer range of about 1200 to 1500 m can probably be explained in terms of the major blocks of woodland cover (see Figure 25).

Summary of Ground Survey

The short-range component of the multivariate variograms reflects reasonably well

the average short-range component of 160 m of the full pixel information from the three channels and NDVI. The average of the long-range component for the full pixel information of 640 m is close to the longer range structure evident in the individual classes for the ground cover classes and the multivariate variogram with 17 classes. The longer range structure of the multivariate variogram with seven classes suggests thet there is an even longer range structure which also emerged from the pixel information of the transects.

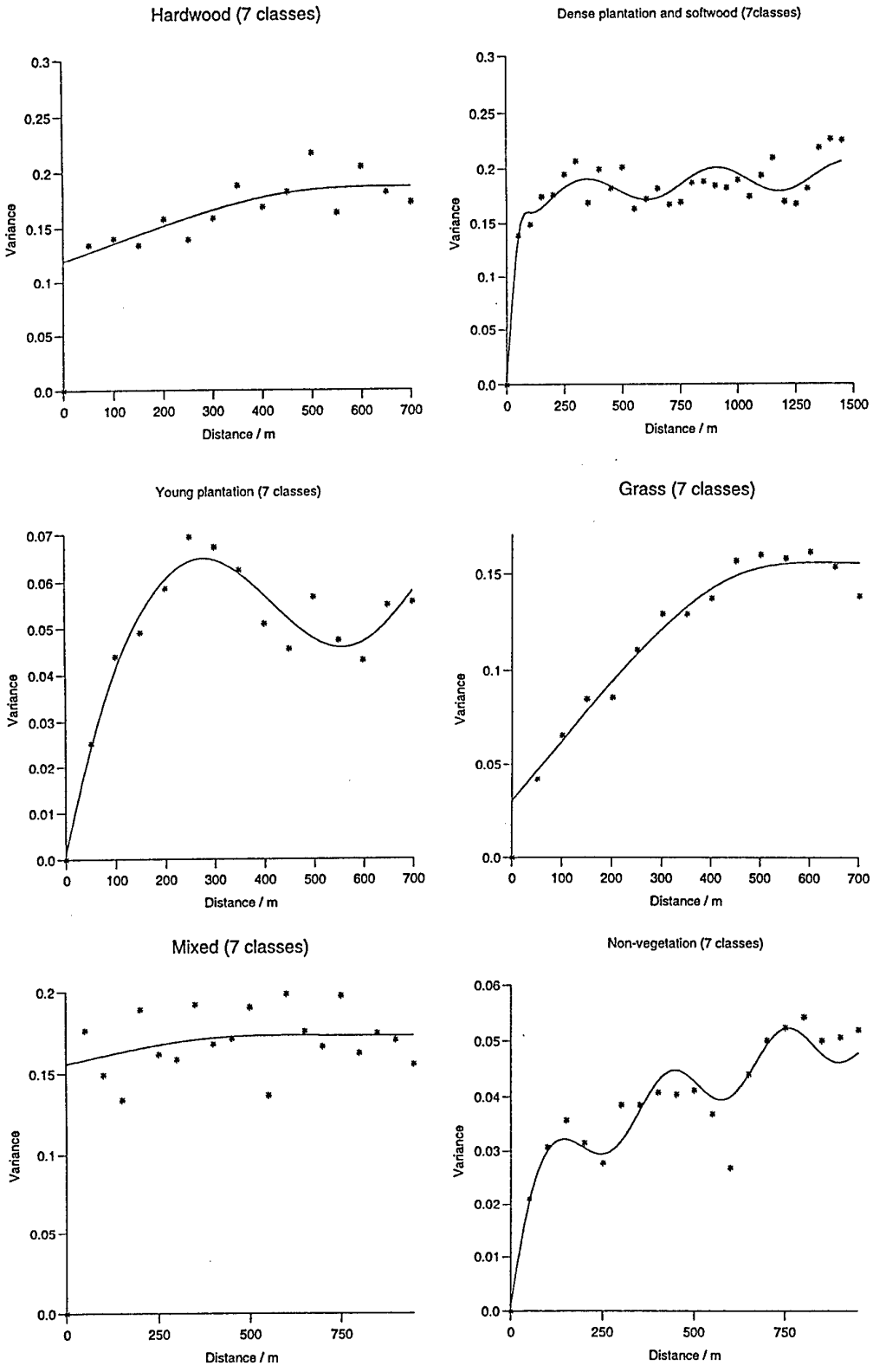

Figure 22: Variograms from transect data of ground cover for seven classes at A. P. Hill. The symbols are the experimental semivariances, and the lines are the fitted models.

Table 6: Variogram model parameters for the ground cover from transect surveys at A. P. Hill with seven classes

		Parameters					
Ground cover	Model	c_0	c_1	c_2	a_1/m	a_2/m	
Multivariate 7 classes	Double sph	2.8825	1.474	2.784	202.0	1582.0	
Hardwood	Circular	0.1193	0.0662		515.5		
Grass	Circular	0.0298	0.1248		480.5		
		Parameters					
	Model	c_0	m	A	B	α	λ/m
Dense plantation	Periodic with power	0	0.1589	–0.0109	–0.0060	0.0581	567.0
Young plantation	Periodic with power	0	0.0370	–0.0112	–0.0076	0.1795	670.0
Industry or road	Periodic with power	0	0.0202	-0.0029	0.0040	0.3167	318.0
Mixed	Pure nugget	Variance					

In the table c_0 is the nugget variance,
c_1 and c_2 are the sill variances of the short- and long-structures, respectively, and
a_1 and a_2 are the distance parameters of the short- and long-structures.
For the periodic with power function m is the gradient and α the exponent of the power term, A and B determine the amplitude and phase, and λ is the wavelength, i.e. the average distance at which repetition in the ground cover occurs.

Table 7: Variogram model parameters of pixel information from SPOT image that coincided with the ground cover transects for A. P. Hill .

		Parameters					
Channel	Model	c_0	m	α	A	B	λ/m
Red	Periodic with power	0	0.99726	0.412	−1.0014	1.1087	496.7
Green	Periodic with power	0	4.6501	0.292	−3.1617	−2.1032	394.7
NIR	Periodic with power	0	80.9018	0.083	−20.3992	−16.5106	503.4
NDVI	Periodic with power	0	0.00135	0.151	−0.00056	−0.00021	520.6

		Parameters		
		c_0	c_1	a_1/m
Red	Spherical	4.45	15.38	1195.0
Green	Exponential	9.88	29.58	1224.6
NIR	Circular	5.55	145.06	219.7
NDVI	Exponential	0	0.00398	330.2

In the table c_0 is the nugget variance,
c_1 is the sill variance of the structure, $a1$ is the distance parameter.
For the periodic with power function m is the gradient and α is the exponent of the power component, A and B determine the amplitude and phase, and λ is the wavelength, i.e. the average distance at which repetition in the ground cover occurs.

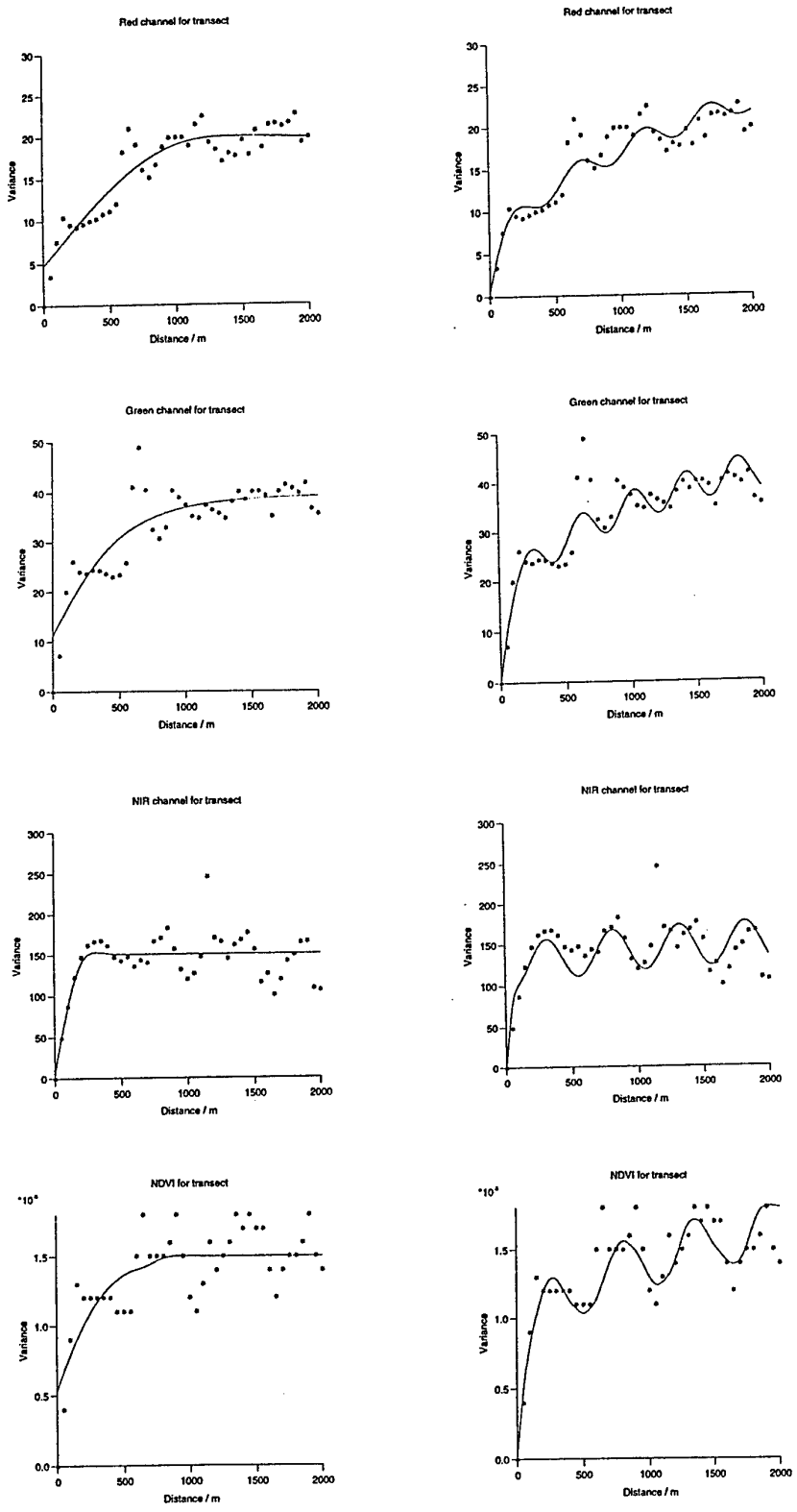

Figure 23: Variograms from pixel information of SPOT image along two transects that coincided with the ground cover survey. The symbols are the experimental semivariances, and the lines are the fitted models.

PART IV: ONE METRE RESOLUTION IMAGERY

FIRST ONE METRE IMAGERY DATA

Information at 1 m resolution was provided for an area within A. P. Hill, but not part of our project area. They were analysed in the first instance to assess the feasibility of handling data at a 1 m resolution. Variograms were computed for the two halves of the area independently and for the entire scene. It was a long narrow flight area with 1.25 million pixels in total. There were four channels and variograms were computed and modelled for each. The results for the entire area are given in Table 8 and Figure 24. They were computed at a 1 m spacing. This means that only short range variation has been identified. All channels show a short range component of variation of about 12 m and a longer one of about 160 m. The latter matches the short range component identified from the SPOT data for our study area.

Table 8: Variogram model parameters for the four channels from the first set of 1 m data for an area at A. P. Hill, but not our study area.

		Variance			Range/m	
Channel	Model	c_0	c_1	c_2	$3a_1$	$3a_2$
1 (Blue) – average	Double exp	0.0	26.89	12.42	12.06	188.25
2 (Green) – average	Double exp	0.0	33.77	11.78	11.07	155.10
3 (Red) – average	Double exp	0.0	63.84	35.82	13.65	212.61
4 (NIR) – average	Double exp	0.0	274.1	112.6	11.88	137.85

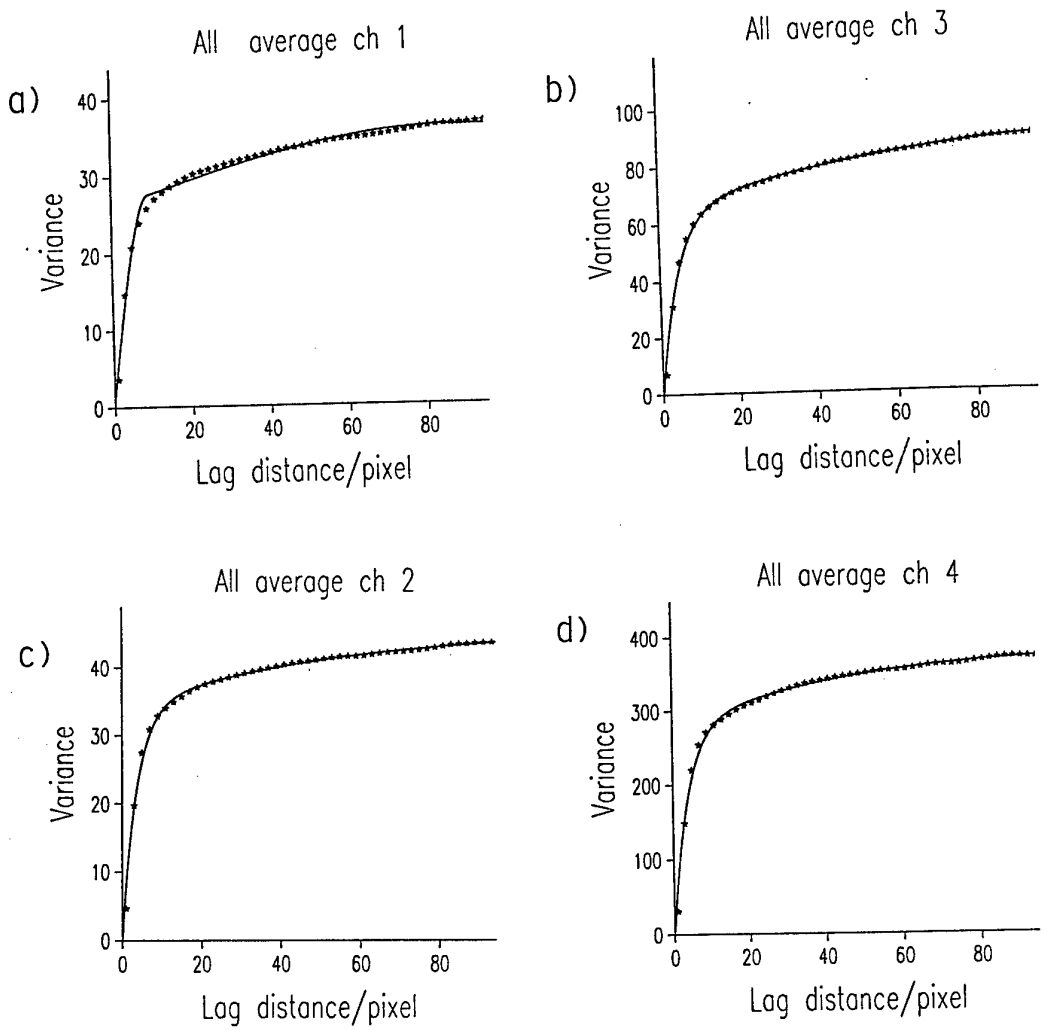

Figure 24: Variograms from the first set of 1 m data for A. P. Hill, not our study area, for four channels: a) Blue, b) Green, c) NIR, d) Red. The symbols are the experimental semivariances, and the lines are the fitted models.

ONE METRE IMAGERY DATA FOR STUDY AREA

The 1 m resolution data for our study area at A. P. Hill, described in part I of this report, are made up from a photo-mosaic, Figure 25. As a result there are some gaps in the overall cover and there are also differences in the intensity of the DNs. The latter is likely to have an effect on the variogram at the longer scales. However, the results show that this has not affected our interpretation of the results unduly. The data were recorded at a height of 1675 m in July 1997. Each pixel covered an area on the ground of 1 m $\times$ 1 m. There are four wavebands: channel 1 – blue, channel 2 – green, channel 3 – near infra red (NIR) and channel 4 – red. Figure 25 is a colour composite of these wavebands from Erdas Imagine.

The data comprise 8 499 498 pixels. This is a large file for geostatistical analysis albeit a modest one for remote sensing applications. As a result of its size it could not be edited as was needed for our purposes. It had already been subdivided at TEC into two subsets: a northern one with 2 610 271 pixels and a southern one with 5 889 224 pixels so that one could be used for validation at some stage. These files could not be edited either. To examine information that comprised pixels from the two subsets a file covering part of both areas was extracted with half a million pixels. These data were then analysed. The northern and southern files were subdivided further: the northern part into 6 subfiles and the southern part into 10 subfiles to facilitate analysis. Each of these contained approximately half a million pixels, and were analysed separately. The results are not included because they do not provide any additional information to that from the first sub-file described below and the selection of 1 m data from the full site.

Subsampling the Data

To explore fully the range of spatial scales that might be present we sampled the data progressively. In this way we can examine variation at several sampling intervals and over different spatial extents to determine the range of scales of spatial variation that are present. Since this also reduced the size of the data considerably we could merge the subfiles for the North and South areas and edit the two main subsets to analyse them. The sub-sampling was as follows:

Sub-sample 1 in 2: A sample of 1 pixel in every 2 along each row and column. This means that one pixel in every four was retained for analysis, giving a 2 m separation between the centroids of each pixel and its nearest neighbour. This

Figure 25: Map of the 1 m data for the study area at A. P. Hill from a colour composite of the four wavebands.

reduced the data by 75%.

Sub-sample 1 in 3: A sample of 1 pixel in every 3 along each row and column, i.e. one pixel in every nine was retained for analysis. This reduced the data by 89%.

Sub-sample 1 in 4: A sample of 1 pixel in every 4 along each row and column, i.e. one pixel in every 16 was retained for analysis. This reduced the data by 93.75%.

Sub-sample 1 in 6: A sample of 1 pixel in every 6 along each row and column, i.e. one pixel in every 36 was retained for analysis. This reduced the data by 97.22%.

Sub-sample 1 in 10: A sample of 1 pixel in every 10 along each row and column, i.e. one pixel in every 100 was retained for analysis. This reduced the data by 99%. Although this might seem to be a huge reduction in data the centroids of the pixels were only 10 m apart and 58 892 remained in the southern area. This is still a large set of data for a reliable geostatistical analysis.

Sub-sample 1 in 15: A sample of 1 pixel in every 15 along each row and column, i.e. one pixel in every 225 was retained for analysis. This reduced the data by 99.6%. Even so 26 174 pixels remained for analysis in the southern area and 37 553 for the whole area.

We decided not to go further at this stage to a sample of 1 pixel in every 400, i.e. 1 in 20 pixels along each row and column because this would have taken us to the resolution of the SPOT image. However, we might do so during the next project. Summary statistics are not given for all of these sub-samples because they are so similar to each other.

Summary statistics

Summary statistics were computed for the 1 in 2 sub-sample for the North and South subsets, and for the entire site. They are given in Table 9. It is evident that channels 2 (green) and 4 (red) are the most skewed as was the case for SPOT data in this area. These wavebands have larger skewness values for the northern subset than for the southern one. After transformation by common logarithms there was still moderate skewness remaining. It is interesting to note that we could not transform the SPOT data for these wavebands satisfactorily and we carried out the analysis on the raw

data in the main. The coefficient of variation (CV) shows that channels 3 and 4 are the most variable, and that channel 1 (blue) is the least variable.

Problems that can arise from computing the variogram with skewed data are more likely where there are few data. Nevertheless we have computed the variograms for some examples from the raw and transformed data.

Table 9: Summary statistics for selected 1 m data and a sub-sample of 1 in 2 from 1 m data of the study area at A. P. Hill.

Channel 1 in 2 selection	Mean	Variance	Standard deviation	Coefficient of variation	Skewness
1 Blue					
North	179.62	194.52	13.95	7.77	3.98
South	181.53	633.29	25.17	13.87	-2.70
All	183.09	268.31	16.38	8.95	2.41
2 Green					
North	55.45	44.26	6.65	11.99	21.84
South	56.85	36.94	6.08	10.70	3.93
All	56.31	35.67	5.97	10.28	11.70
3 NIR					
North	96.30	371.63	19.28	18.98	1.66
South	100.46	321.81	17.94	17.86	0.68
All	99.63	337.33	18.37	18.44	1.31
4 Red					
North	50.00	52.71	7.26	14.52	6.75
South	53.07	196.71	14.03	26.44	6.75
All	51.56	125.81	11.22	21.76	8.67
1 m selection					
1 Blue	178.76	190.11	13.79	7.71	4.17
2 Green	55.81	152.56	12.35	22.15	14.97
3 NIR	98.26	363.75	19.07	19.41	-0.52
4 Red	49.88	102.48	10.12	20.29	14.54

Variography of One Meter Resolution Imagery

Variograms were computed for all of the data files along the rows and down the columns for a total of 100 lags. The average variogram of the rows and columns was also computed in each case. The variograms were computed using a FORTRAN 77 program written for the previous project by R. Webster, and they were modelled using Genstat (Genstat5 Committee, 1993). For sub-samples 1 to 3 above all of the variograms were modelled, but for the other data the average variogram only was modelled. A range of models (see earlier in the report) was tried, but in general a double (nested) exponential model fitted the best in the least squares sense and a double spherical one less frequently. In some cases the variogram showed a sharp increase after it had levelled to its sill. The usual interpretation of this is that there is trend in the data; for these data the most likely explanation is the variation in the intensity of the DNs from image processing. It occurs mainly where the variogram has been computed over longer distances, for instance the 1 in 10 and 1 in 15 samplings. Where this has occurred the variogram has been modelled to a shorter lag distance.

To summarise before going into the detail of the variogram models and the figures – the variograms along the rows invariably had longer distance parameters than those for the columns for both the long and the short range components. This accorded with our findings for the SPOT image, and for the pixels extracted from the transects described earlier in the report. In addition the distance parameters were longer for the southern part of the image. However, as noted for the SPOT images, Figures 1 to 3, there is no marked anisotropy except possibly for the areas with large reflectance values.

There is some variation in the range of spatial dependence defined by the models fitted to the variograms. However, certain values stand out forcibly which is evident in the tables for each waveband.

Results

Channel 1 – Blue

Figure 26 shows a map of the pixel values for channel 1 using Imagine. They were given as a grey scale only, but it is possible to see some of the variation present. The light areas are the roads, paths, buildings and areas of hardstanding. The intermediate shades are grass and water bodies, and the rest appears to be woodland or shrub. There is little contrast in this waveband, except for the light areas, which is what we should expect given the coefficient of variation.

Variograms were computed for all of the subsets. The results of these analyses are summarised in Table 10. The experimental variograms (symbols) and the fitted models (solid line) are shown in Figures 27 to 29. The scales of variation that emerge as significant are: approximately 20 to 30 m for the short-range component, and 200 to 250 m, and 300 to 400 m for the long-range component.

Figure 26: Map of the Blue waveband values from the 1 m data for the study area.

Table 10: Variogram model parameters for channel 1 (Blue) for the 1 m data from the study area at A. P. Hill.

Sub-sample	Model	Variance			Range/m	
		c_0	c_1	c_2	$a_1/3a_1$	$a_2/3a_2$
North						
1 in 2 – row	Double exp	0.0	52.98	137.60	27.78	401.58
1 in 2 – column	Double exp	0.0	38.37	123.90	14.77	213.59
1 in 2 – average	Double exp	0.0	43.18	126.40	19.08	264.18
1 in 3 – average	Double exp	0.0	43.31	126.20	18.90	262.30
1 in 4 – average	Double exp	0.0	52.75	123.80	25.92	323.20
1 in 6 – average	Double exp	19.33	41.04	118.20	50.22	340.40
1 in 10 – average	Double exp	24.69	32.69	120.00	52.50	324.60
1 in 15 – average	Double sph	49.03	66.10	58.70	130.10	408.80
South						
1 in 2 – row	Double exp	0.0	81.25	192.70	35.88	435.24
1 in 2 – column	Double exp	0.0	79.63	174.40	26.94	401.04
1 in 2 – average	Double exp	0.0	80.24	183.90	31.20	418.30
1 in 3 – average	Double exp	0.0	76.13	173.60	31.86	436.40
1 in 4 – average	Double exp	9.11	71.49	172.30	39.72	467.30
1 in 6 – average	Double exp	12.65	66.61	171.20	39.78	451.62
1 in 10 – average	Double exp	0.0	95.81	163.60	48.60	548.78
1 in 15 – average	Double sph	0.0	81.07	169.00	37.35	464.40
Site						
Selected area 1 m – average	Double exp	0.0	80.24	183.90	15.6	239.00
1 in 2 – average	Double exp	0.0	64.13	162.70	27.00	351.40
1 in 4 – average	Double exp	0.0	69.98	164.50	30.96	401.50
1 in 10 – average	Double sph	42.63	68.66	116.60	74.50	365.82
1 in 15 – average	Double sph	48.79	63.93	110.20	86.66	359.87

a_1 and a_2 are the ranges for the double spherical model, and
$3a_1$ and $3a_2$ are the ranges for the double exponential model.

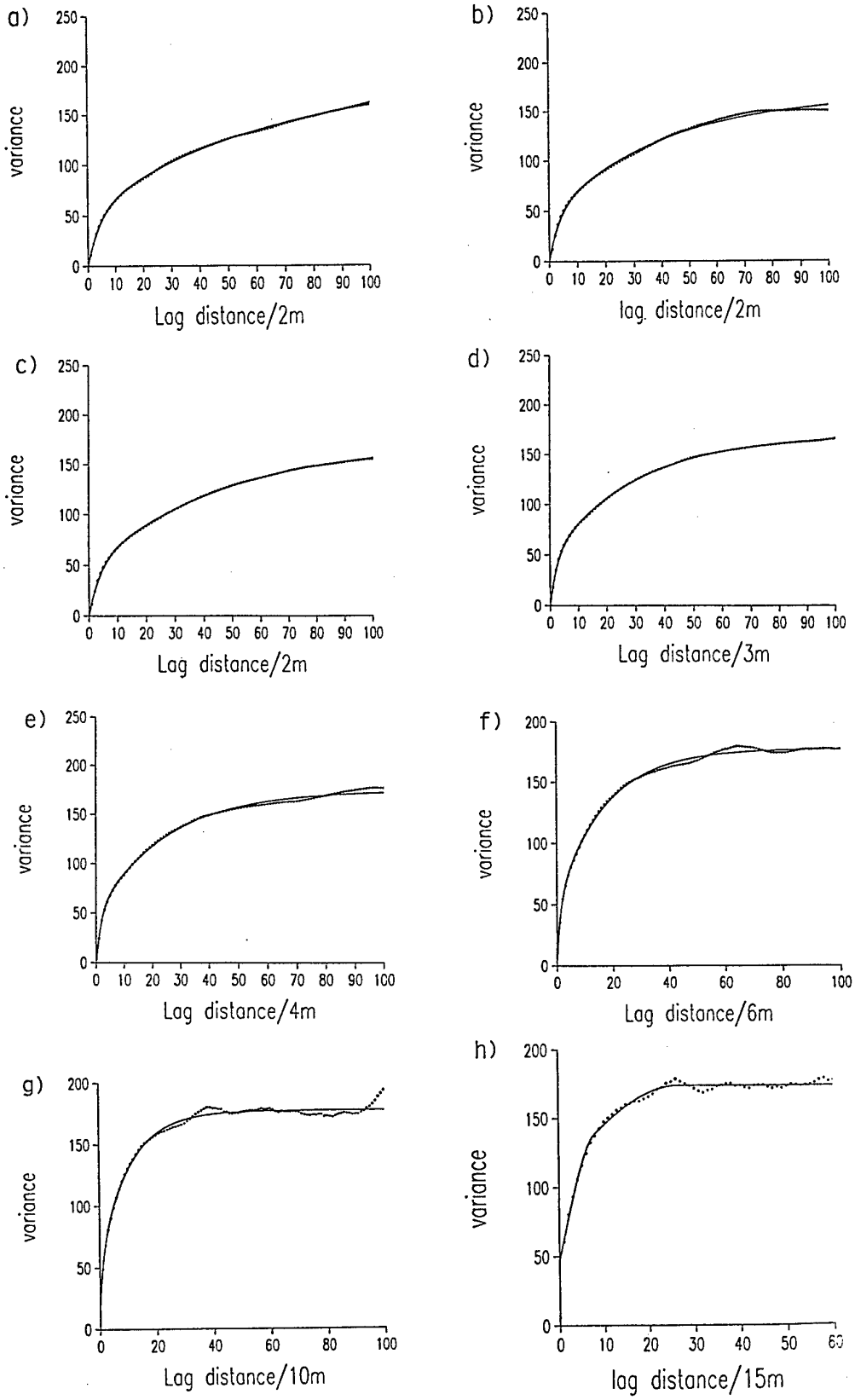

Figure 27: Variograms of channel 1 (Blue) from 1 m data for the N part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average, f) 1 in 6 average, g) 1 in 10 average, h) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

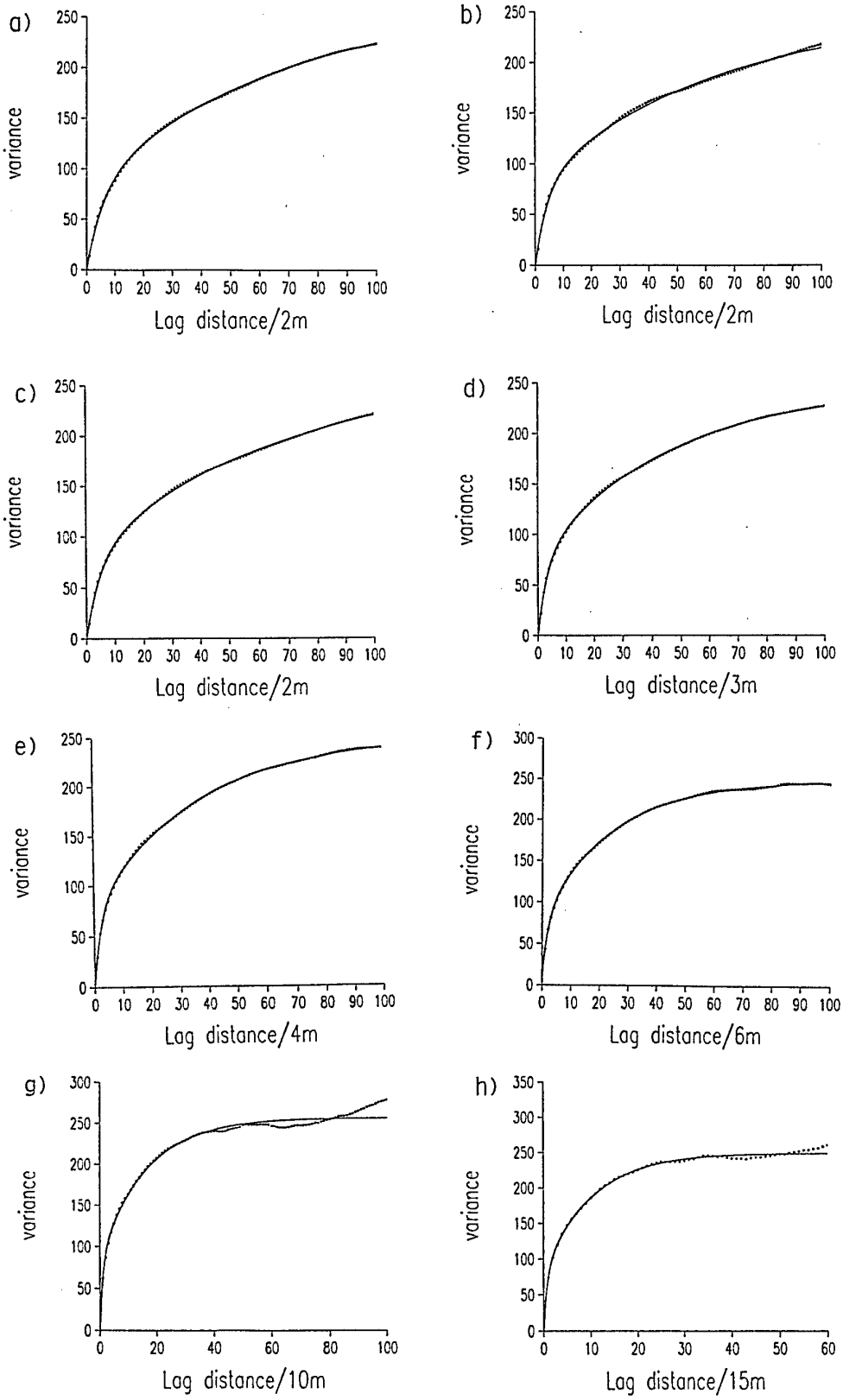

Figure 28: Variograms of channel 1 (Blue) from 1 m data for the S part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average, f) 1 in 6 average, g) 1 in 10 average, h) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

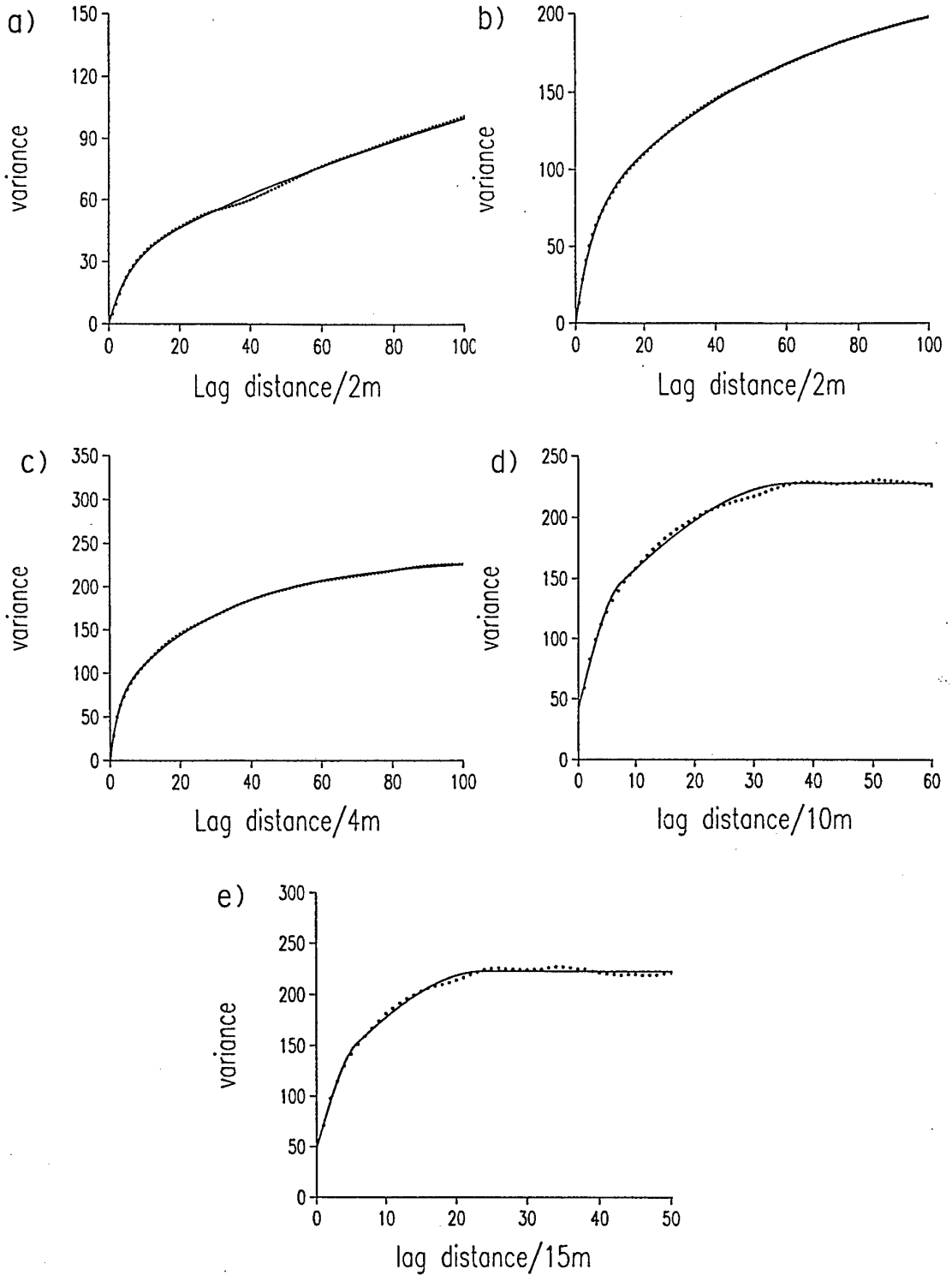

Figure 29: Variograms of channel 1 (Blue) from 1 m data for the study area: a) 1 m selection average, b) 1 in 2 average, c) 1 in 4 average, d) 1 in 10 average, e) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

Channel 2 – Green

Figure 30 shows a map of the pixel values for channel 2 as before. More contrast is evident in this waveband. The areas of hardstanding, roads, paths and buildings remain light. Grass has intermediate shades, and the water bodies stand out as the darkest patches in this map. There is also more variation in the wooded areas than in the blue waveband.

Variograms were computed for all of the subsets. The results of these analyses are summarised in Table 11 and Figures 31 to 34. Some of the variograms of this channel have a distinctly wavy form, especially those along the rows (Figure 31). Periodicity was also evident in the variograms of the ground cover along the transects (Figures 20 and 22), and in the pixel data from these transects (Figure 23). The scales of variation that emerge as significant for the raw data are: 15 to 20 m and 50 to 70 m for the short-range component, and 100 m, 400 to 500 m, and about 1 km for the long-range component. For the data transformed to logarithms the short-range component is about 100 m and the long range one about 1.5 km to 2 km for the northern subset.

Figure 30: Map of the Green waveband values from the 1 m data for the study area, A. P. Hill.

Table 11: Variogram model parameters for channel 2 (Green).

Sub-sample	Model	Variance			Range/m	
		c_0	c_1	c_2	$a_1/3a_1$	$a_2/3a_2$
North						
1 in 2 – row	Spherical	3.68	28.88		16.8	
1 in 2 – column	Double exp	0.0	18.98	27.09	11.58	171.96
1 in 2 – average	Double sph	3.41	23.26	11.04	14.96	103.32
1 in 3 – average	Double sph	4.37	21.50	10.08	15.24	96.45
1 in 4 – average	Double exp	0.0	52.75	123.80	15.84	86.36
1 in 6 – average	Double exp	19.33	41.04	118.20	13.26	107.88
1 in 6 – $\log_{10}$	Double sph	0.00038	0.00037	0.00262	105.90	684.60
1 in 10 – average	Double exp	24.69	32.69	120.00	72.40	175.00
1 in 10 – $\log_{10}$	Double exp	0.00031	0.00046	0.00078	137.10	1606.8
1 in 15 – average	Double sph	49.03	66.10	58.70	71.17	909.60
1 in 15 – $\log_{10}$	Double exp	0.00024	0.00040	0.00032	152.96	2250.00
South						
1 in 2 – row	Double exp	0.0	9.78	16.65	26.64	246.66
1 in 2 – column	Double exp	0.77	8.10	16.35	20.46	297.84
1 in 2 – average	Double exp	2.54	9.67	16.20	47.28	450.00
1 in 3 – average	Double exp	2.15	9.52	14.31	51.93	495.00
1 in 4 – average	Double exp	2.61	9.69	14.48	57.00	545.30
1 in 6 – average	Double exp	3.01	8.23	14.64	50.58	292.80
1 in 10 – average	Double exp	2.75	7.59	14.73	56.70	434.70
1 in 10 – $\log_{10}$	Double exp	0.0	0.00053	0.00071	51.00	539.40
1 in 15 – average	Double sph	6.13	8.21	10.67	83.70	413.00
1 in 15 – $\log_{10}$	Double sph	0.00086	0.00131	0.00172	74.60	386.50
Site						
Selected area 1 m – average	Double exp	0.0	108.90	37.08	141.30	239.00
Selected area 1 m – $\log_{10}$	Double exp	0.0	0.00113	0.00126	13.17	281.40
1 in 2 – average	Double exp	0.0	12.82	13.74	16.92	179.34
1 in 4 – average	Double exp	0.0	14.62	12.82	20.52	216.96
1 in 10 – average	Double exp	8.44	14.19	5.79	111.39	811.50
1 in 15 – average	Double exp	6.10	10.86	9.81	90.00	900.00

a_1 and a_2 are the ranges for the double spherical model, and
$3a_1$ and $3a_2$ are the ranges for the double exponential model.

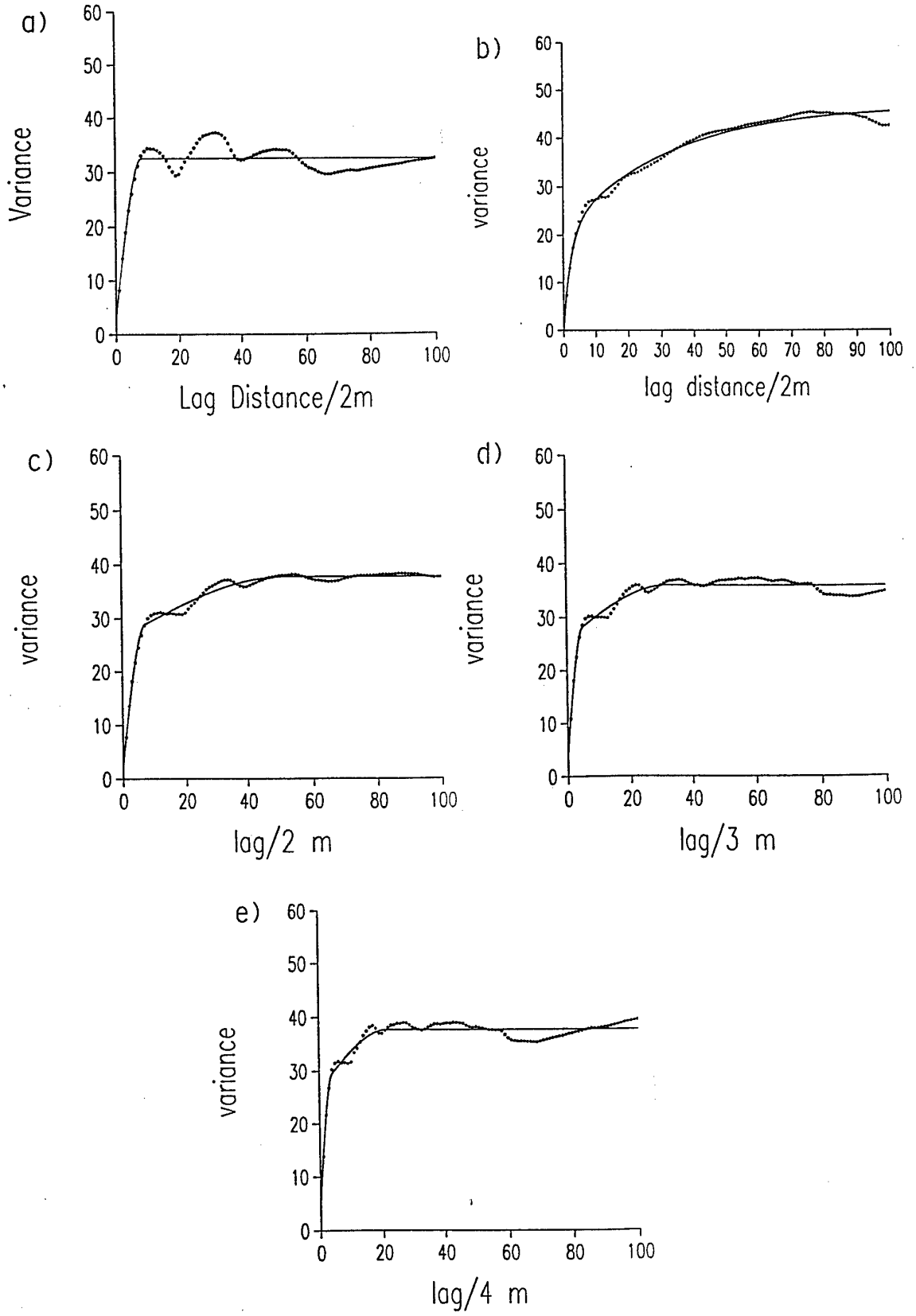

Figure 31: Variograms of channel 2 (Green) from 1 m data for the N part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average. The symbols are the experimental semivariances, and the lines are the fitted models.

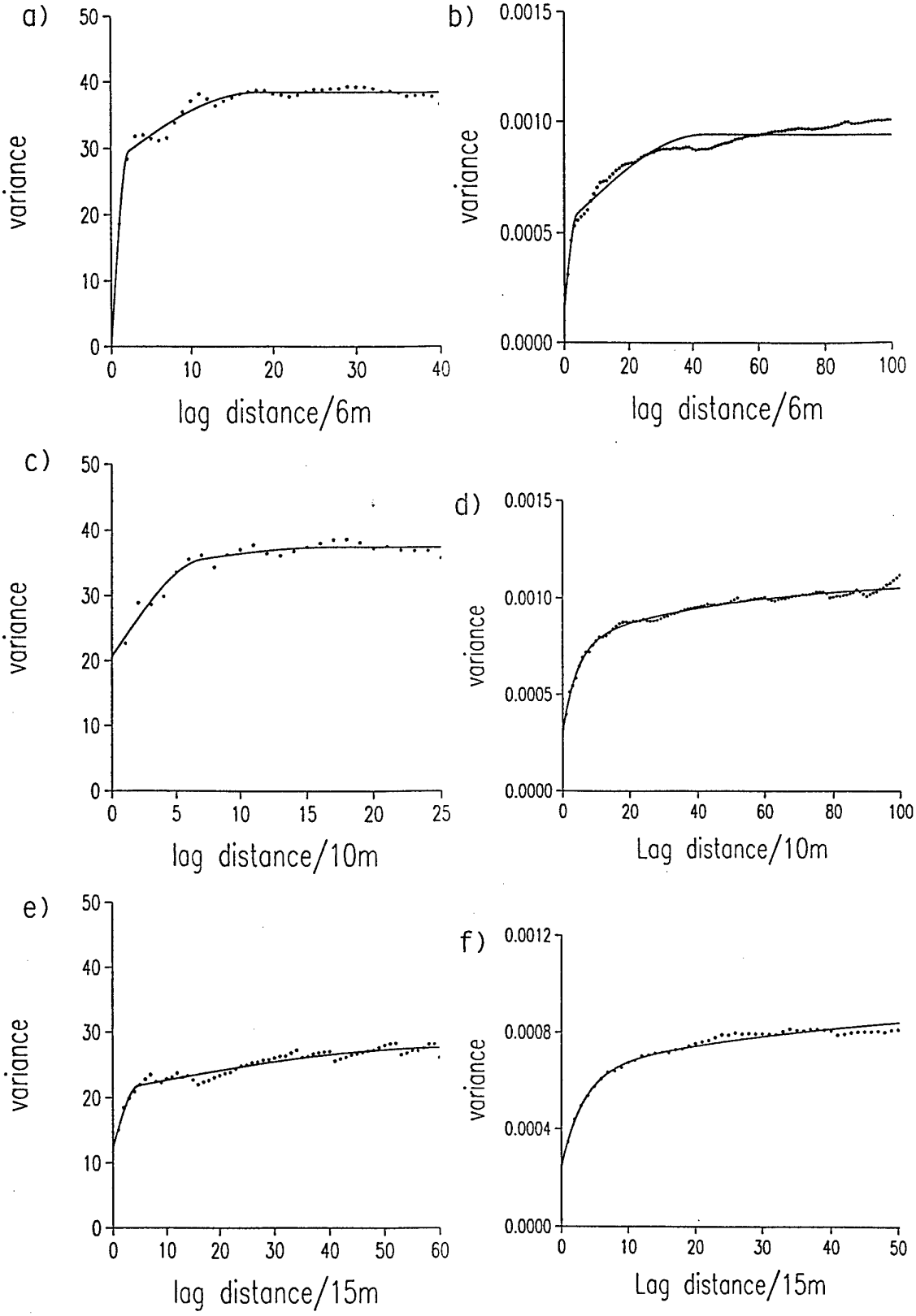

Figure 32: Variograms of channel 2 (Green) from 1 m data for the N part of study area: a) 1 in 6 average, b) 1 in 6 average log transformed, c) 1 in 10 average, d) 1 in 10 average log transformed, e) 1 in 15 average, f) 1 in 15 average log transformed. The symbols are the experimental semivariances, and the lines are the fitted models.

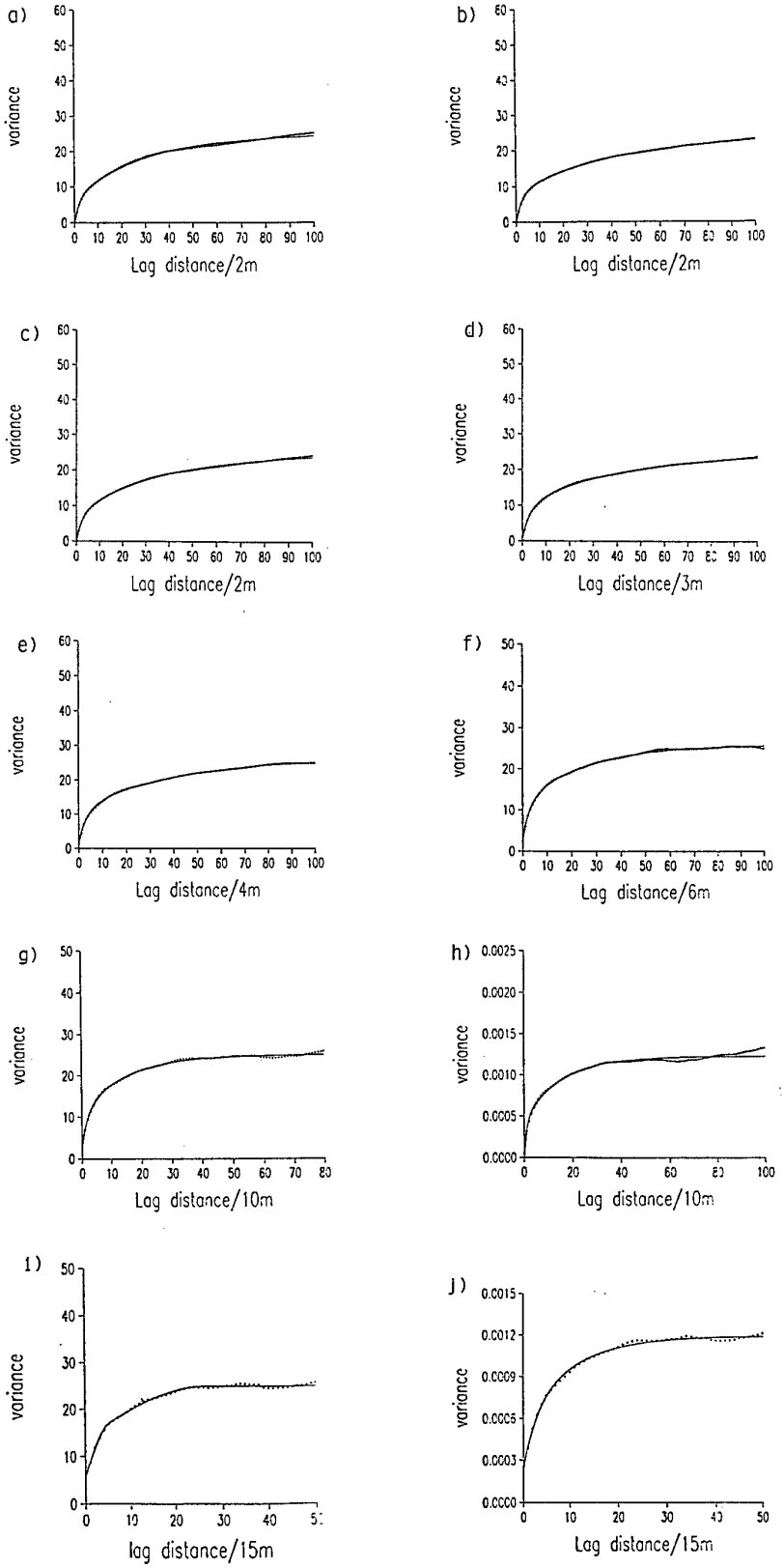

Figure 33: Variograms of channel 2 (Green) from 1 m data for the S part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average. f) 1 in 6 average, g) 1 in 10 average, h) 1 in 10 average log transformed, i) 1 in 15 average, j) 1 in 15 average log transformed. The symbols are the experimental semivariances, and the lines are the fitted models.

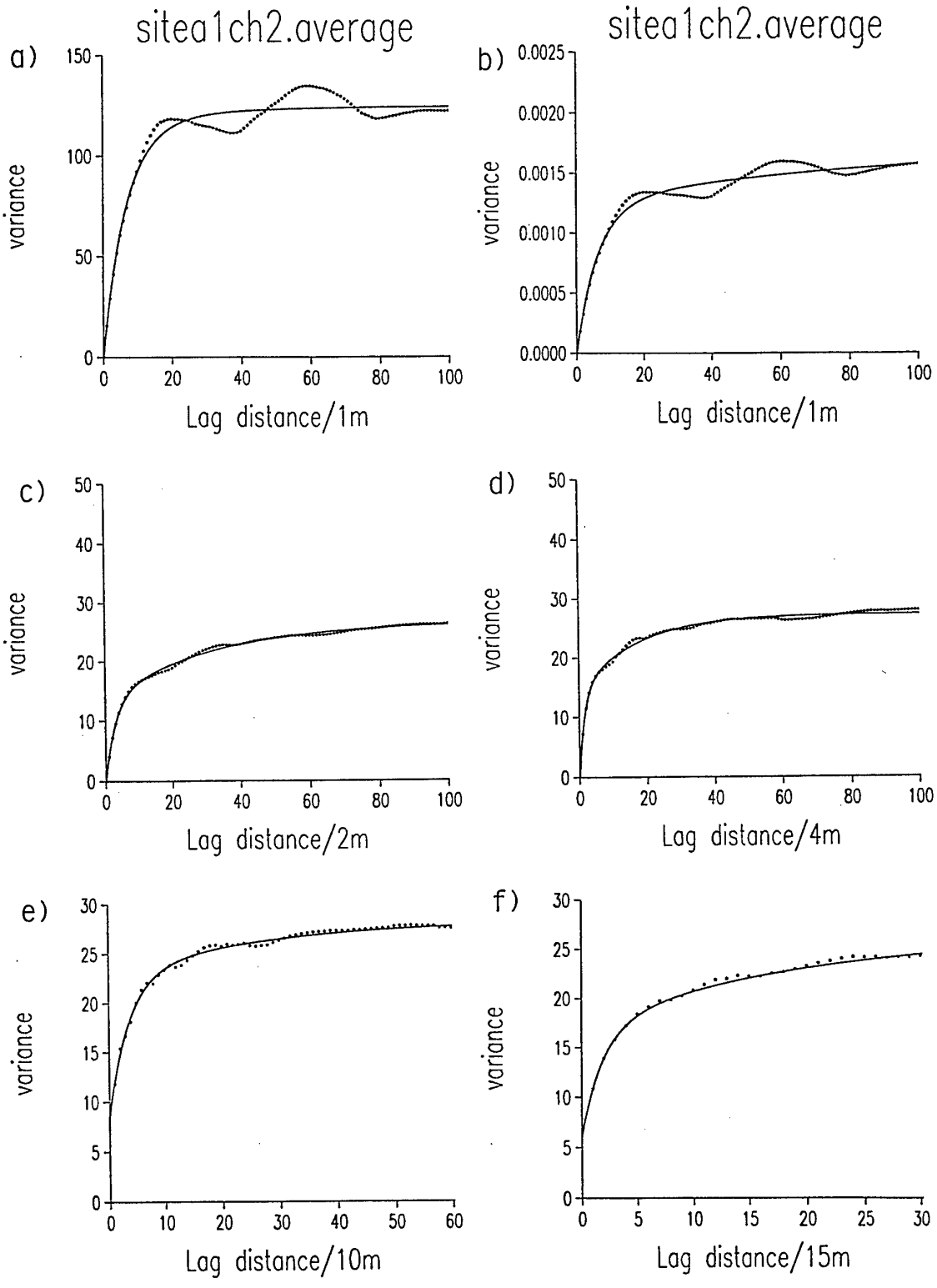

Figure 34: Variograms of channel 2 (Green) from 1 m data for the study area: a) 1 m selection average, b) 1 m selection average $\log_{10}$, c) 1 in 2 average, d) 1 in 4 average, e) 1 in 10 average, f) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

Figure 34: Variograms of channel 2 (Green) from 1 m data for the study area: a) 1 m selection average, b) 1 in 2 average, c) 1 in 4 average, d) 1 in 10 average, e) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

Channel 3 – Near Infra Red

A map of the pixels for this waveband is shown in Figure 35. It shows considerable variation, especially at the intermediate scales, and appears to contain much information. For this waveband the buildings, roads and water bodies are dark, and the grass is light. There is evidence of considerable variation in the areas of woodland. The map of NIR values for the SPOT image, Figure 3, also shows considerable variation overall.

Variograms were computed as before for all of the subsets. The results are summarised in Table 12 and Figures 36 to 38. This waveband has the smallest skewness values, therefore, these variograms should be the most reliable. The scales of variation that emerge as significant are: 25 to 30 m for the short-range component, as for the blue waveband, and about 200 m, for the long-range component.

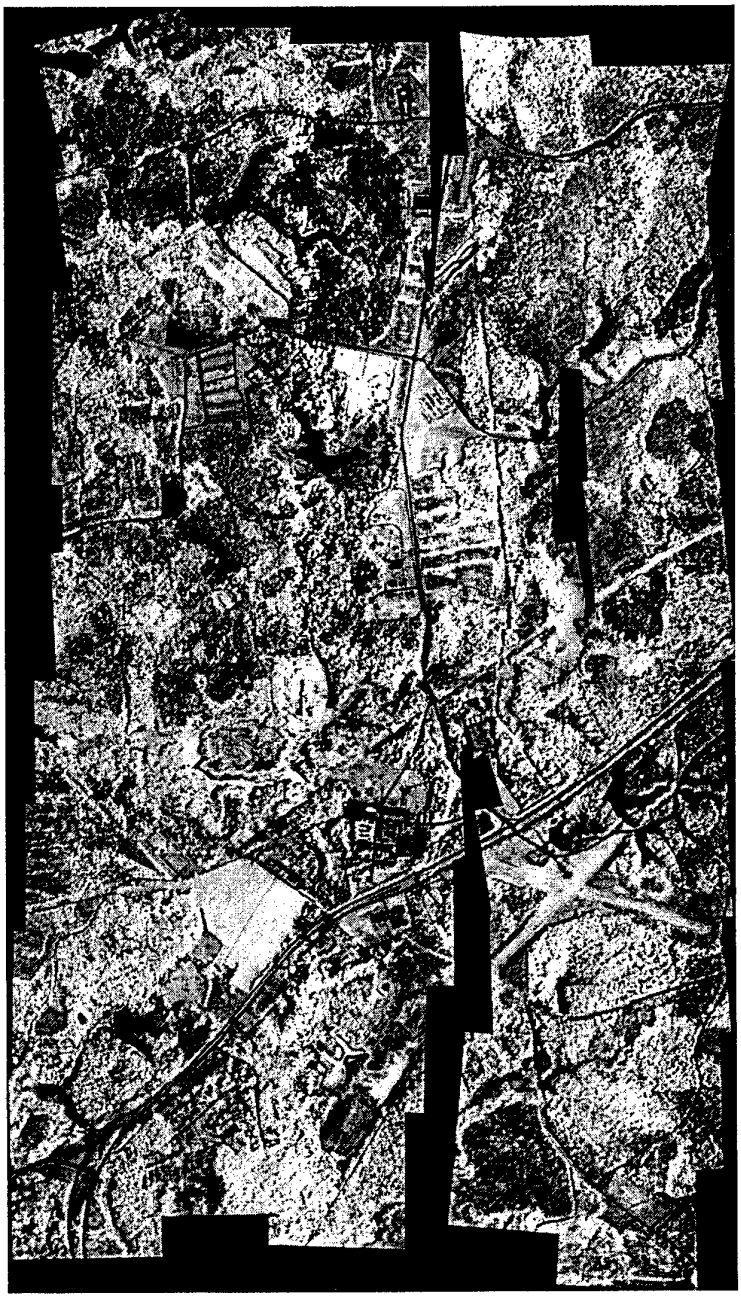

Figure 35: Map of the NIR waveband values from the 1 m data for the study area, A. P. Hill.

Table 12: Variogram model parameters for channel 3 (NIR).

Sub-sample	Model	Variance			Range/m	
		c_0	c_1	c_2	$a_1/3a_1$	$a_2/3a_2$
North						
1 in 2 – row	Double sph	0.0	184.40	111.90	17.06	106.54
1 in 2 – column	Double exp	0.0	170.40	234.10	32.10	322.68
1 in 2 – average	Double exp	0.0	174.50	169.40	26.82	216.24
1 in 3 – average	Double exp	0.0	165.80	174.80	24.84	190.44
1 in 4 – average	Double exp	0.0	143.90	189.10	27.72	154.20
1 in 6 – average	Double exp	0.0	163.50	174.50	24.30	191.70
1 in 10 – average	Double exp	0.0	157.80	184.90	20.70	177.30
1 in 15 – average	Double sph	140.60	83.21	100.60	72.00	214.05
South						
1 in 2 – row	Double sph	0.0	183.20	82.62	21.82	152.68
1 in 2 – column	Double sph	0.0	155.00	108.00	25.40	148.24
1 in 2 – average	Double sph	0.0	169.10	95.10	23.38	150.14
1 in 3 – average	Double exp	0.0	186.60	112.50	32.76	270.27
1 in 4 – average	Double exp	0.0	188.30	112.20	33.12	283.20
1 in 6 – average	Double exp	0.0	183.50	115.10	31.68	261.36
1 in 10 – average	Double exp	36.46	160.40	103.80	40.80	302.70
1 in 15 – average	Double exp	76.52	119.69	101.00	52.20	323.10
Site						
Selected area 1 m – average	Double exp	0.0	90.93	211.80	21.36	97.08
1 in 2 – average	Double exp	0.0	180.20	133.70	30.36	238.98
1 in 4 – average	Double exp	0.0	187.90	130.80	31.68	272.40
1 in 10 – average	Double exp	0.0	189.70	130.90	30.00	278.55
1 in 15 – average	Double sph	130.00	103.80	78.88	75.02	345.47

a_1 and a_2 are the ranges for the double spherical model, and
$3a_1$ and $3a_2$ are the ranges for the double exponential model.

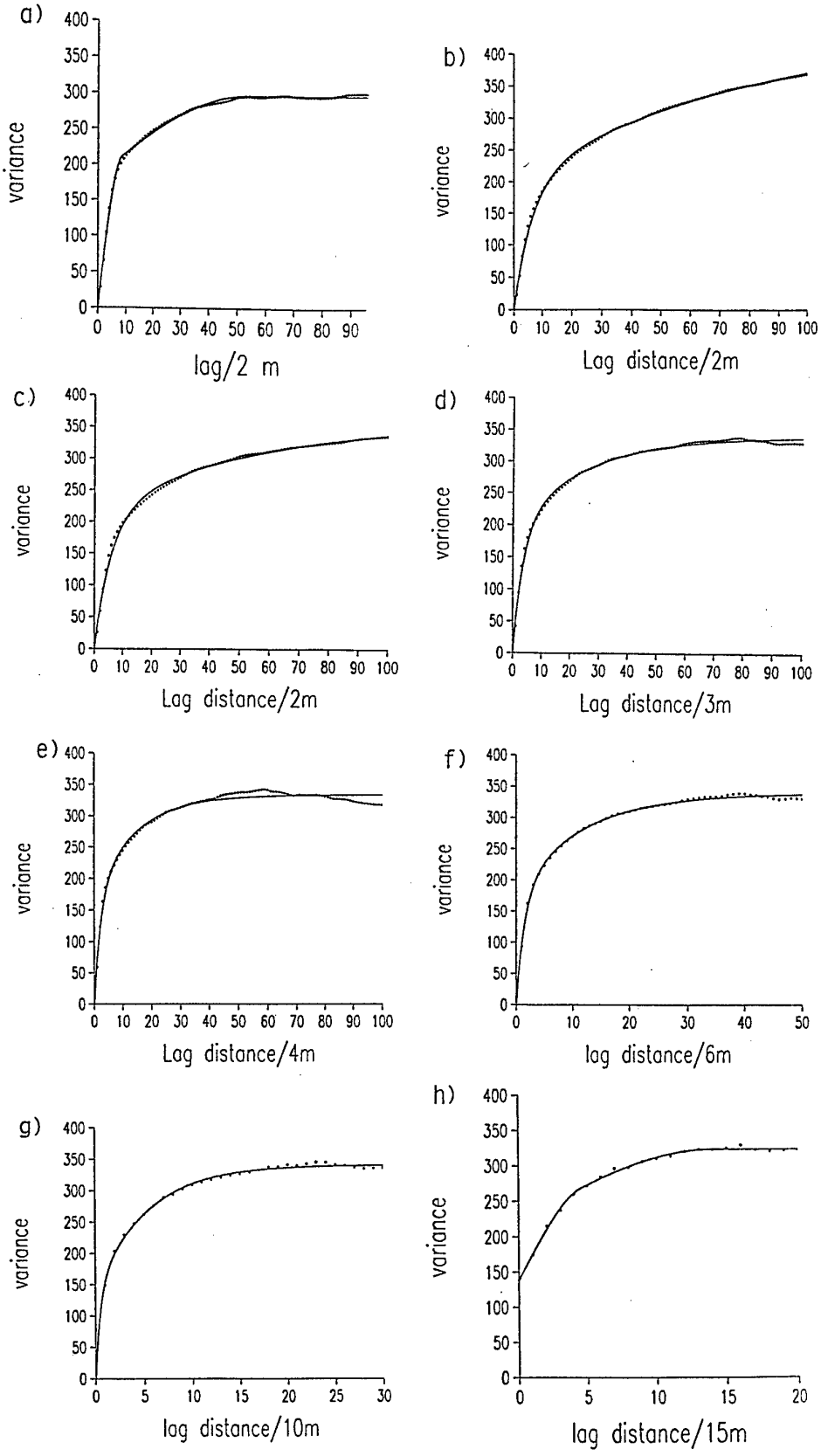

Figure 36: Variograms of channel 3 (NIR) from 1 m data for the N part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average, f) 1 in 6 average, g) 1 in 10 average, h) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

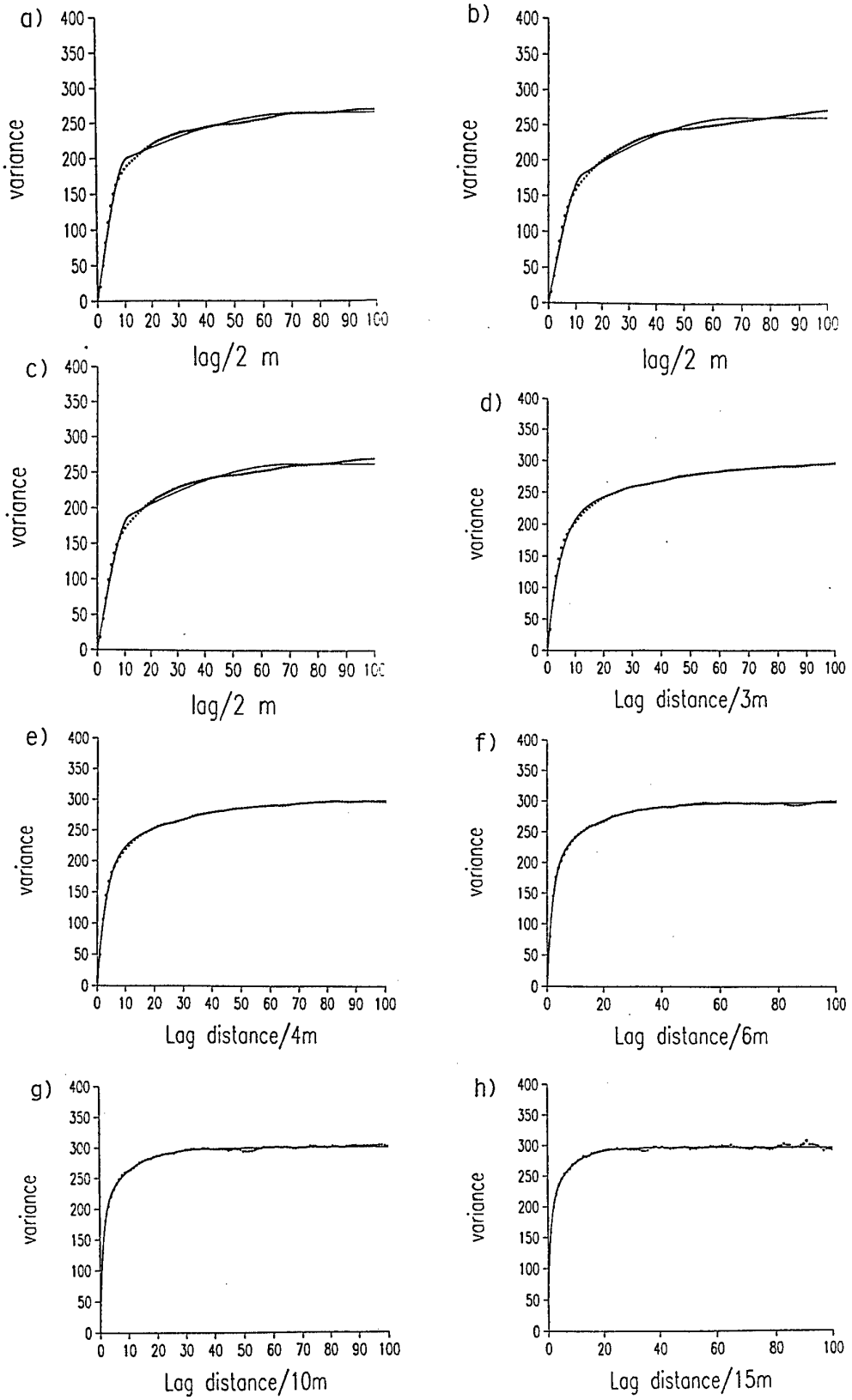

Figure 37: Variograms of channel 3 (NIR) from 1 m data for the S part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average, f) 1 in 6 average, g) 1 in 10 average, h) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

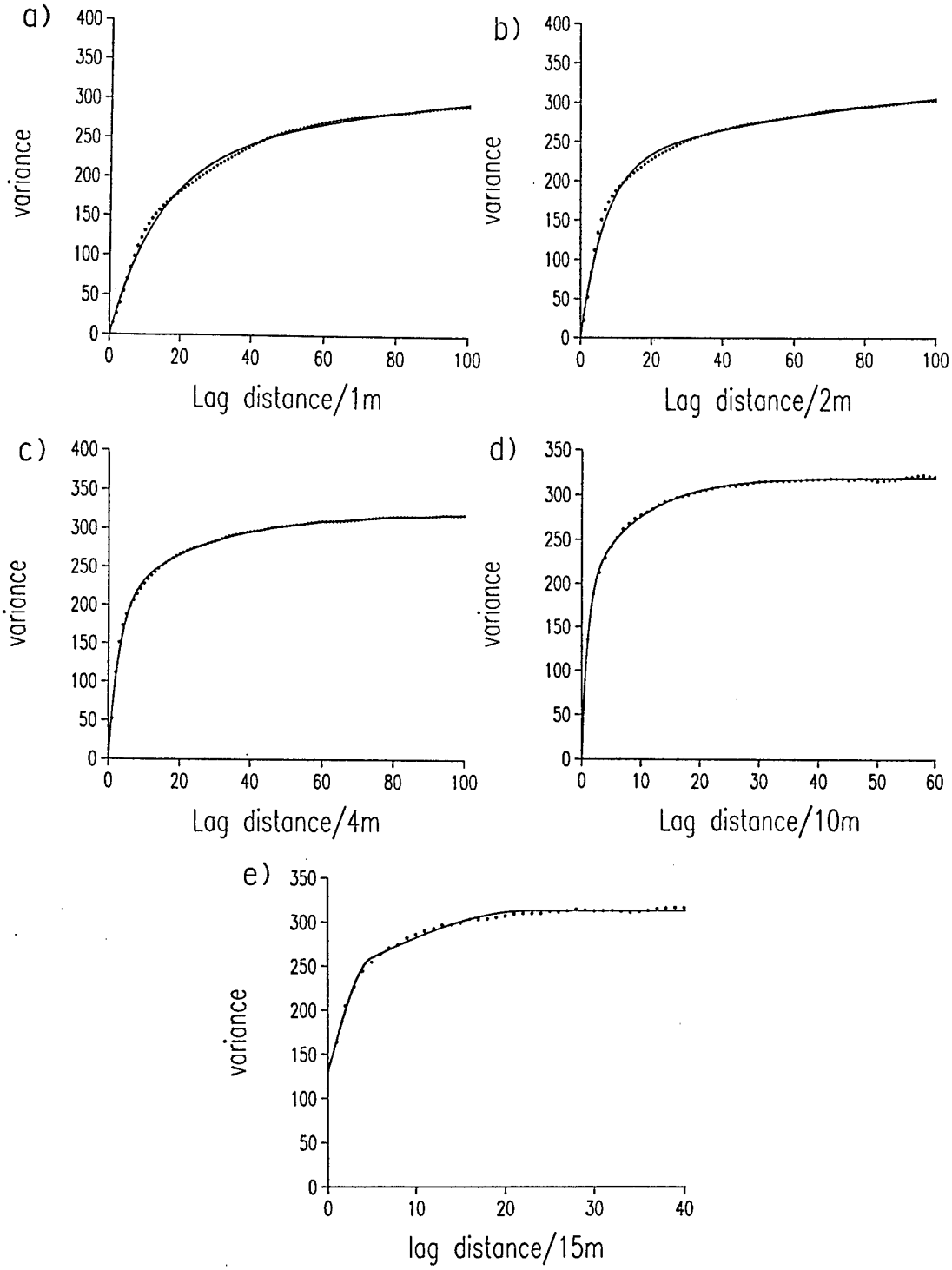

Figure 38: Variograms of channel 3 (NIR) from 1 m data for the study area: a) 1 m selection average, b) 1 in 2 average, c) 1 in 4 average, d) 1 in 10 average, e) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

Channel 4 – Red

A map of the pixels for this waveband is shown in Figure 39. It does not appear to be as variable as the map for NIR, but the values have the largest coefficient of variation. They are also the most skewed. The shades are similar to those in the blue waveband, but more variation is evident in the wooded areas. The water bodies, however, are not evident.

Variograms were computed for all of the subsets. The results of these analyses are summarised in Table 13 and Figures 40 to 44. The scales of variation that emerge as significant for the raw data and that transformed to logarithms are: 15 to 20 m and 75 m for the short-range component, and 150 m, 300 m, 700 to 900 m, and about 1.5 km for the long-range component.

Figure 39: Map of the Red waveband values from the 1 m data for the study area, A. P. Hill.

Table 13: Variogram model parameters for channel 4 (Red).

Sub-sample	Model	Variance			Range/m	
		c_0	c_1	c_2	$a_1/3a_1$	$a_2/3a_2$
North						
1 in 2 – row	Double exp	3.62	27.24	17.04	21.06	535.20
1 in 2 – column	Double exp	0.0	22.31	27.56	11.28	213.42
1 in 2 – average	Double exp	1.94	24.04	21.03	15.42	247.02
1 in 3 – average	Double exp	2.81	22.61	21.28	1.71	25.22
1 in 4 – average	Double exp	9.41	21.19	19.89	27.00	372.24
1 in 6 – average	Double exp	16.42	20.19	20.03	61.02	912.24
1 in 10 – average	Double exp	14.07	19.89	22.96	35.10	505.20
1 in 10 – $\log_{10}$	Double sph	0.00080	0.00057	0.00078	100.50	398.00
1 in 15 – average	Double exp	23.65	14.63	18.96	126.90	1381.95
1 in 15 – $\log_{10}$	Double exp	0.00074	0.00081	0.00064	204.62	890.00
South						
1 in 2 – row	Double exp	0.0	22.86	99.04	7.98	99.00
1 in 2 – column	Double exp	10.857	45.26	63.87	27.06	201.42
1 in 2 – average	Double exp	0.0	35.26	83.94	12.18	122.52
1 in 3 – average	Double exp	0.0	51.58	57.47	27.09	177.93
1 in 4 – average	Double exp	16.78	74.77	49.36	75.48	1704.00
1 in 6 – average	Double exp	18.04	69.16	32.85	72.00	703.80
1 in 10 – average	Double exp	2.75	7.59	14.73	79.80	825.90
1 in 10 – $\log_{10}$	Double exp	0.0	0.00260	0.00179	82.22	933.30
1 in 15 – average	Double exp	14.36	67.37	35.88	66.15	796.50
1 in 15 – $\log_{10}$	Double exp	0.00025	0.000265	0.00063	135.00	450.00
Site						
Selected area 1 m – average	Double exp	0.0	17.99	112.90	15.12	66.90
Selected area 1 m – $\log_{10}$	Double exp	0.0	0.00125	0.00050	19.77	281.80
1 in 2 – average	Double exp	0.0	29.51	57.82	13.38	122.70
1 in 4 – average	Double exp	16.56	57.58	27.16	73.44	889.20
1 in 10 – average	Double exp	22.71	49.91	25.23	74.28	425.70
1 in 15 – average	Double exp	18.49	47.93	26.29	67.19	526.50

a_1 and a_2 are the ranges for the double spherical model, and
$3a_1$ and $3a_2$ are the ranges for the double exponential model.

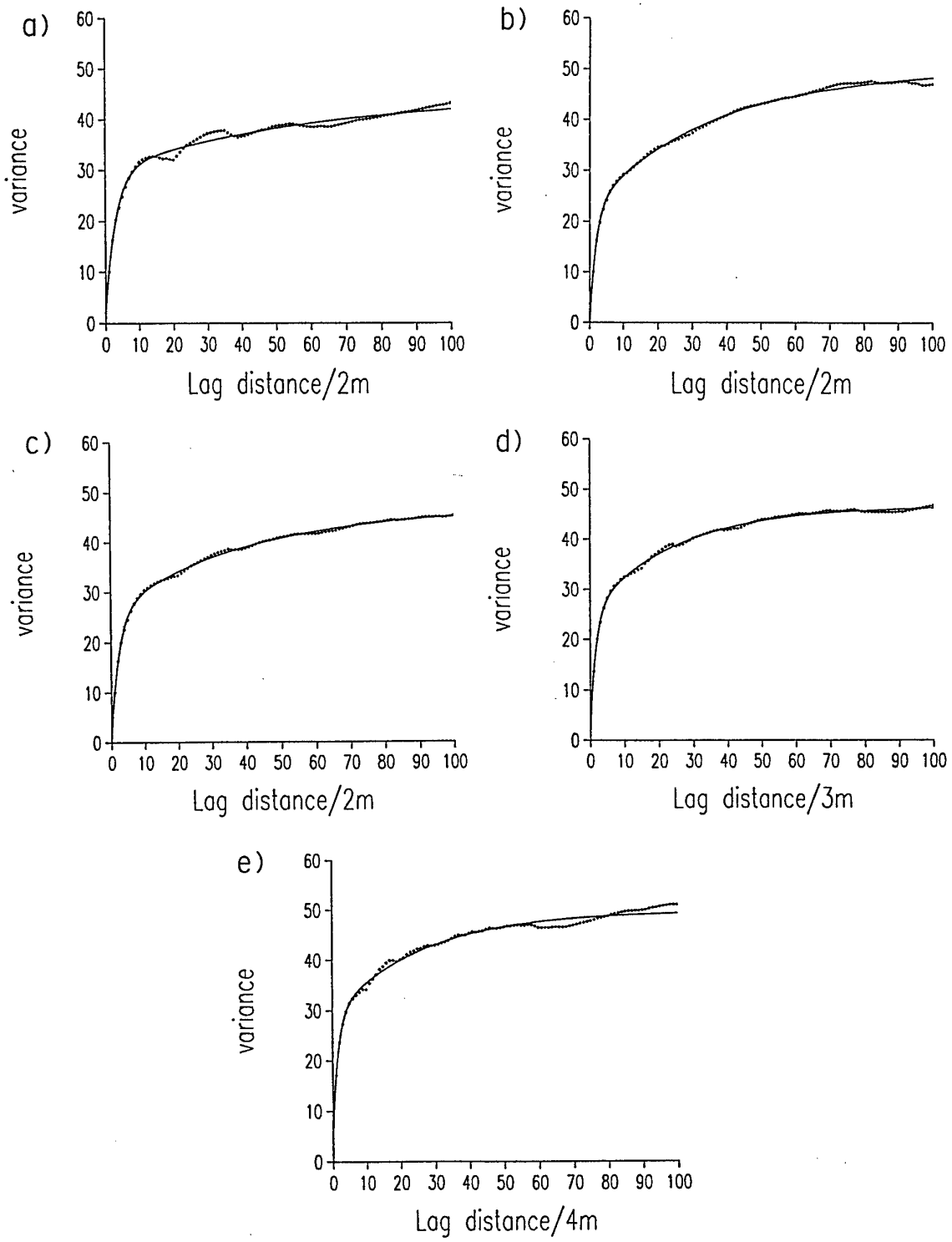

Figure 40: Variograms of channel 4 (Red) from 1 m data for the N part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average. The symbols are the experimental semivariances, and the lines are the fitted models.

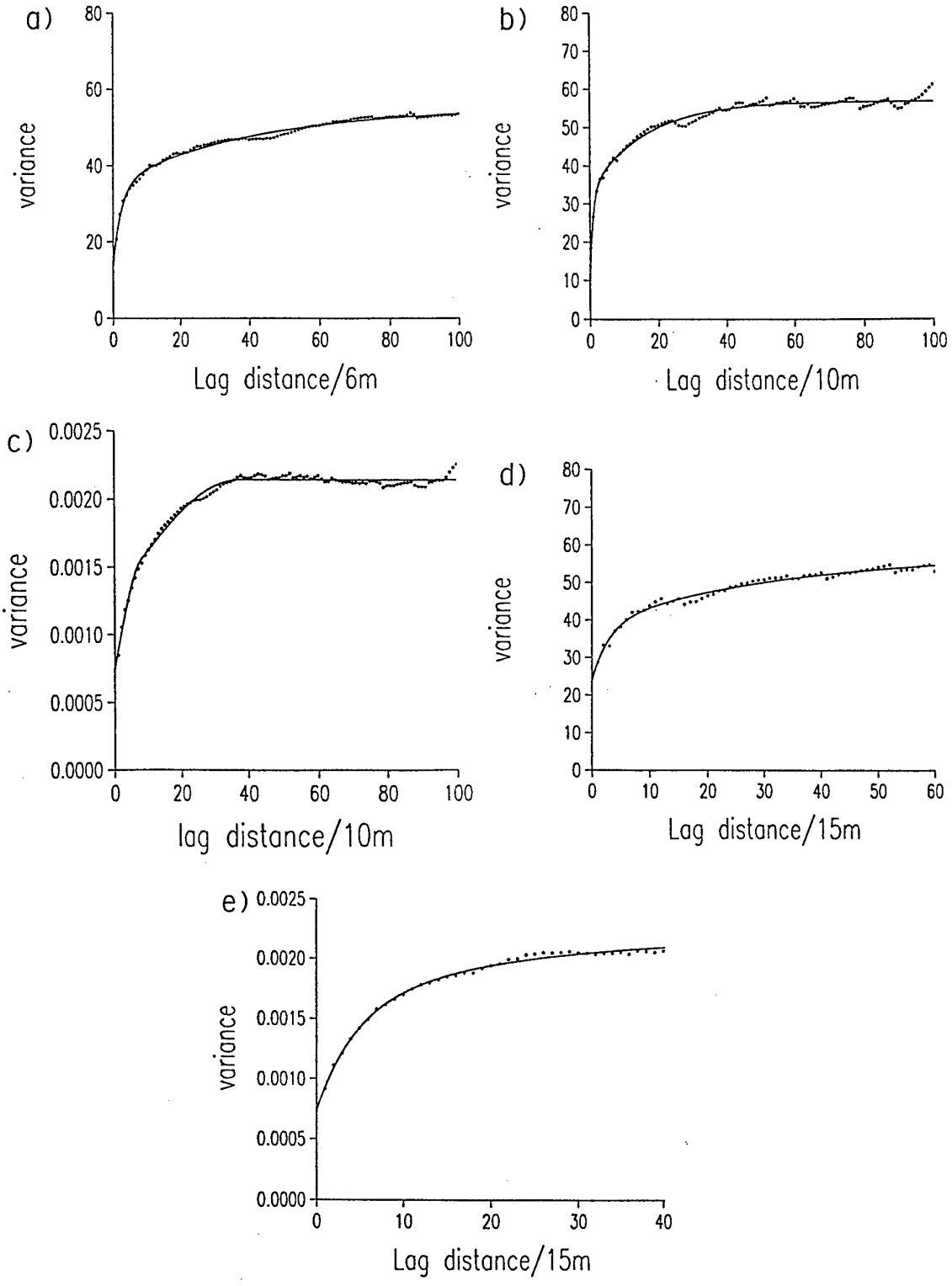

Figure 41: Variograms of channel 4 (Red) from 1 m data for the N part of study area: a) 1 in 6 average, b) 1 in 10 average, c) 1 in 10 average log transformed, d) 1 in 15 average e) 1 in 15 average log transformed. The symbols are the experimental semivariances, and the lines are the fitted models.

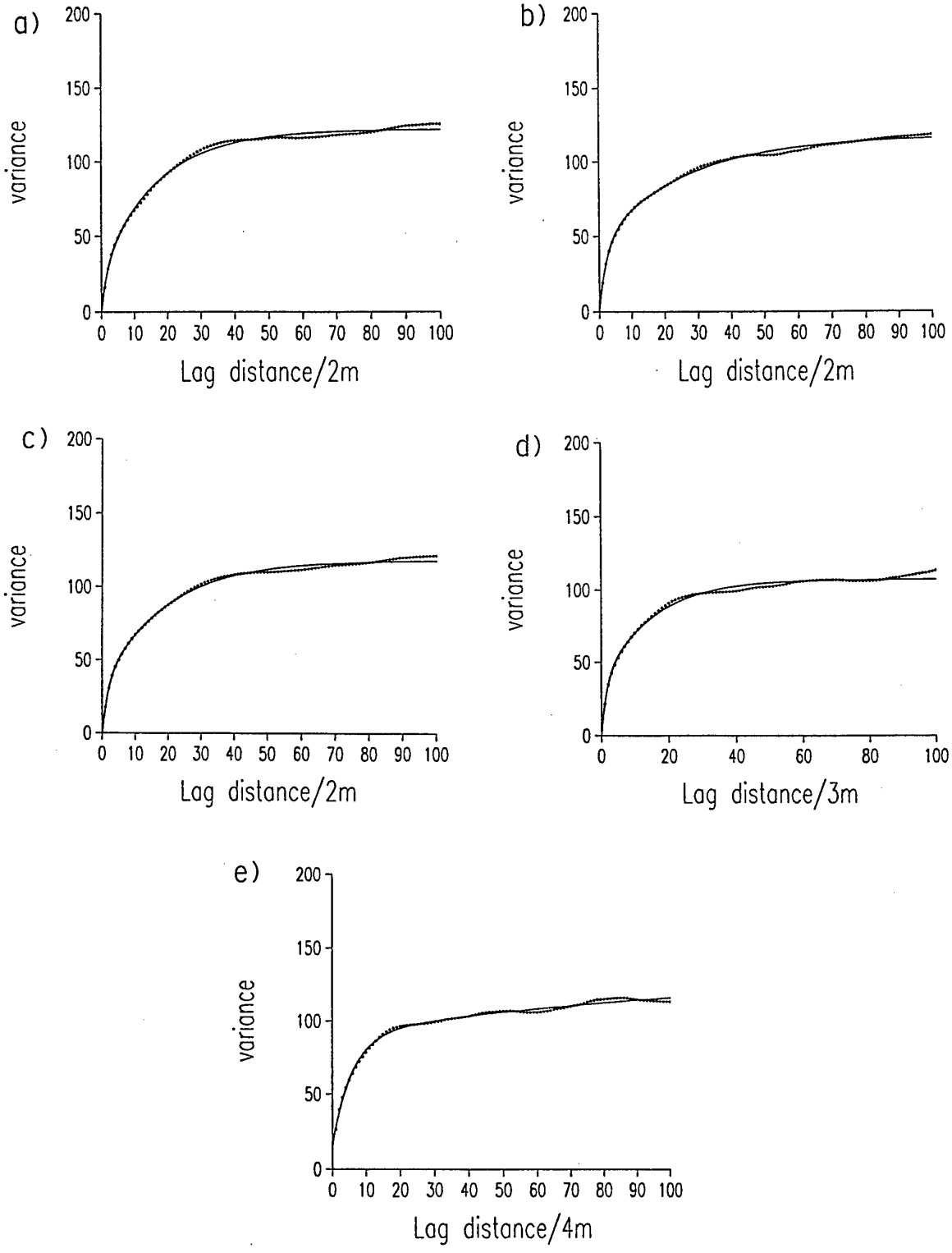

Figure 42: Variograms of channel 4 (Red) from 1 m data for the S part of study area: a) 1 in 2 rows, b) 1 in 2 columns, c) 1 in 2 average, d) 1 in 3 average, e) 1 in 4 average. The symbols are the experimental semivariances, and the lines are the fitted models.

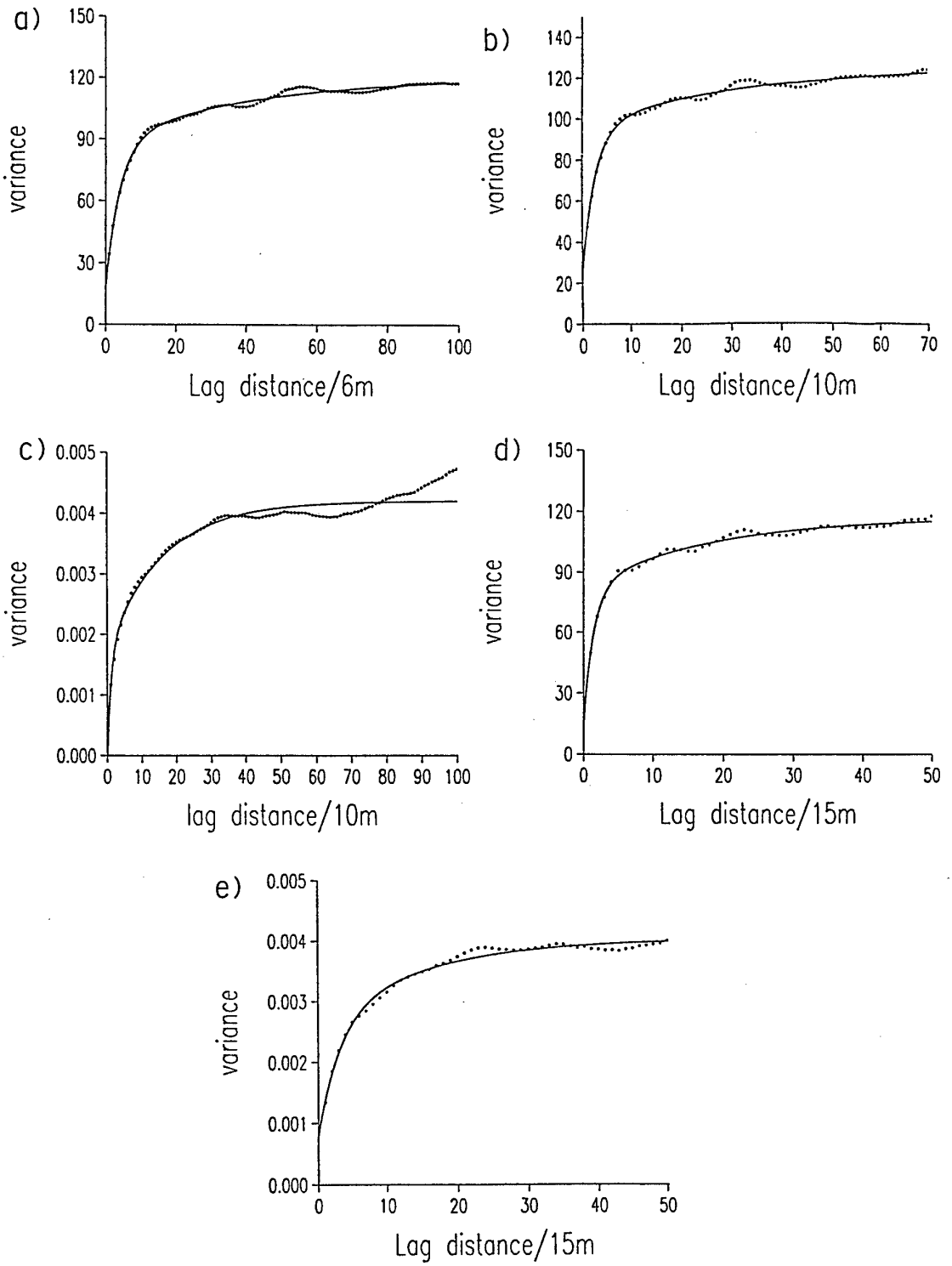

Figure 43: Variograms of channel 4 (Red) from 1 m data for the S part of study area: a) 1 in 6 average, b) 1 in 10 average, c) 1 in 10 average log transformed, d) 1 in 15 average, e) 1 in 15 average log transformed, The symbols are the experimental semivariances, and the lines are the fitted models.

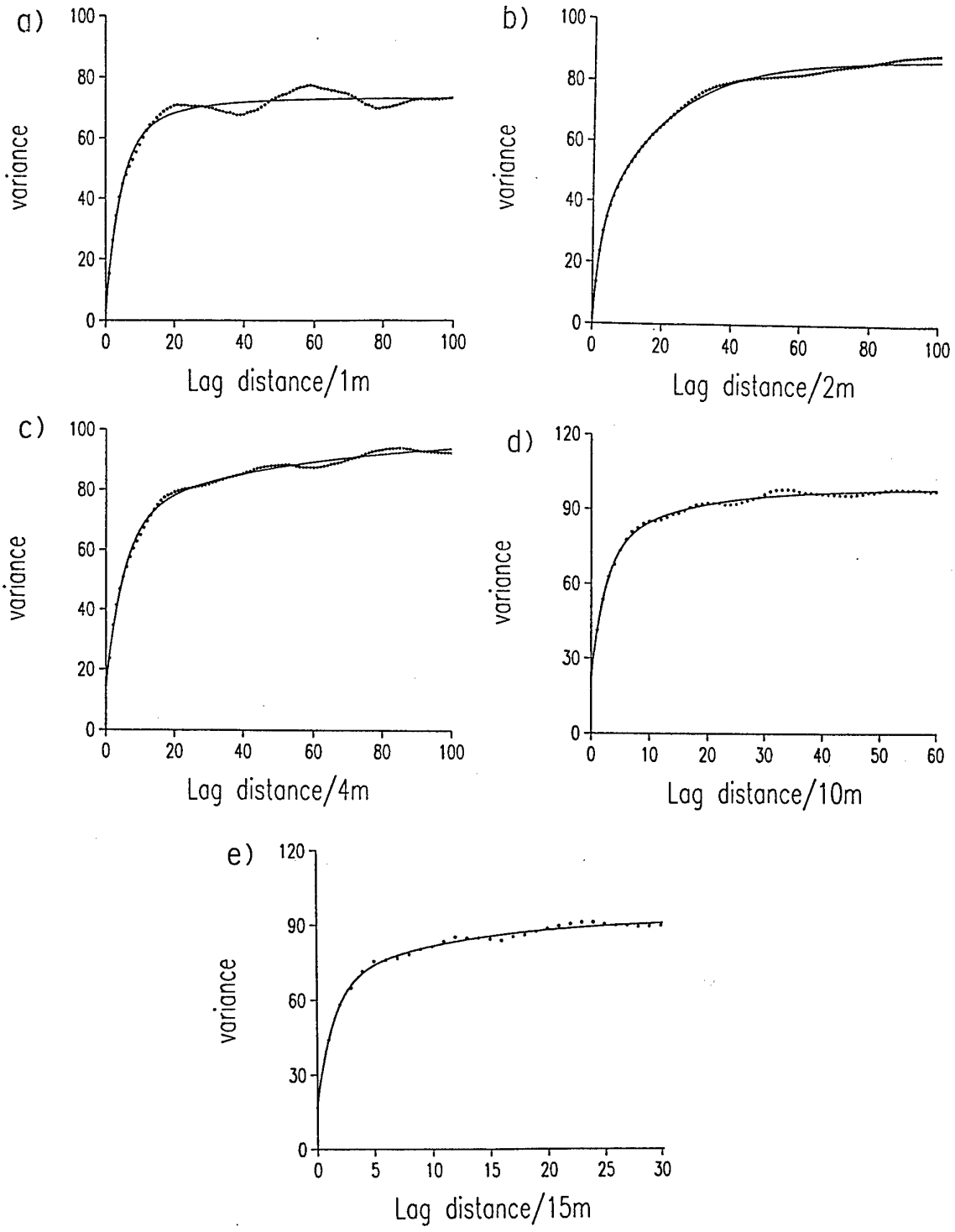

Figure 44: Variograms of channel 4 (Red) from 1 m data for the study area: a) 1 m selection average, b) 1 in 2 average, c) 1 in 4 average, d) 1 in 10 average, e) 1 in 15 average. The symbols are the experimental semivariances, and the lines are the fitted models.

SUMMARY AND CONCLUSIONS OF GEOSTATISTICAL ANALYSES

The variograms of the DNs in the three channels of the SPOT imagery and that of the NDVI derived from them embody spatial variation on two distinct scales, one of 150 m to 200 m and the other 600 m to 700 m. They match the scales in the multivariate variogram of the ground cover classes closely. Kriging analysis of the DNs and NDVI show the pattern in the image at the two spatial scales clearly. The short-range component seems to represent the local physiography, and the long-range one the main types of ground cover.

We have had to develop means of handling very large sets of data for analysis by subdividing them into smaller files, resampling those to reduce their size and then merging them again. There seems to be no more information in the 1-m data than in the 1 in 2 sub-sample, i.e. alternate pixels in each row and column retained. In general the results from the different samplings differed only with the very sparse samplings of 1 in 10 and 1 in 15 of each row and column, which we expected. There are some differences between the wavebands, in particular the range of the longer range component in the near infra red is less than that of the green and red. There are differences between the northern and southern parts of the image that need to be explored further: the South has longer scales of variation than the North.

There seem to be important contributions to the total spatial variation in the 1-m imagery at distances of 20 m to 30 m, 100 m to 200 m, 500 m to 700 m, and about 1 km. The smallest distance appears as the 'pock-marks' in Figure 25, which might be caused by shadows from the tree crowns. This contribution is unlikely to have any practical value on the ground. It coincides closely with the pixel size of the SPOT image, which does not resolve it therefore. The 100 m to 200 m scale of variation corresponds with the short-range variation in the SPOT image. The 500 m to 700 m scale relates to the longer range component in the SPOT data. The 1-km range was evident in the NDVI from SPOT.

The results suggest strongly that there are consistent scales of spatial variation evident in the SPOT and 1-m imagery, and in the ground cover data. Little additional information on significant spatial scales appears to have been obtained from the 1-m imagery. More spatial scales emerge from the 1-m data, and they seem to relate closely to those for the different wavebands of the SPOT data. The 20 m $\times$ 20 m pixel size of the SPOT seems to be the ideal resolution for examining the variation in ground cover and for removing the local effects of light and height. The 1-m imagery is presumably better for identifying buildings and roads, but provides more

information than necessary for survey of ground cover and for management. There is much redundancy in the 1-m data, though we might gain more insight into the different spatial scales from the data when we compare them with the more detailed information on ground cover an elevation that is now available.

The next stages of the work will examine the relation between the pixel values and the ground cover class for the SPOT data and that at 1-m resolution. Since there seems to be a strong physiographic control on the variation we shall also do a coregionalization analysis of the pixel information with the elevation data. In addition we shall explore the two parts of the area for which we have SPOT and 1-m data in more detail using kriging analysis and wavelets.

REFERENCES

Bourgault, G. & Marcotte, D. 1993. Spatial filtering under the linear coregionalization model. In: A. Soares (ed.) *Geostatistics Troiá '92*, pp. 237-248.

Genstat 5 Committee (1993) *Genstat 5 Release 3 Reference Manual*, Oxford University Press, Oxford.

Wen, R. & Sinding-Larsen, R. 1997. Image filtering by factorial kriging - sensitive analysis and applications to Gloria side-scan sonar images. *Mathematical Geology*, **29**, 433-468.

PART V: AIDE MEMOIRE FOR SPATIAL ANALYSIS OF IMAGE AND GROUND DATA

Introduction

Raw data from scanners aboard aircraft and satellites are commonly processed at the receiving stations to correct them geometrically and create files of pixel values of relative intensity of light in several separate wavebands. The pixels are typically square and are assumed to be disjoint and contiguous. Their coordinates may be given or implied. The files are then passed to analysts in branches of earth science and biology whose concern is to make sense of the data, to map them, and perhaps to predict variables of interest. The scientists will often have data from ground surveys, usually the coordinates given, whether they are on regular grids or transects or irregularly scattered. They may wish to analyse these data similarly.

This document summarizes the steps that a scientist should take in a geostatistical analysis of such data, beginning with error detection, summary statistics, exploratory data analysis, the variogram and its modelling, kriging, and mapping. A summary of some of the important Genstat commands are given on Page 86. Appendix III contains a set of example programs for all of the analyses discussed in the Aide Memoir.

Notation

The following notation is used in this aide. The geographic coordinates of the centres of the pixels, sampling points, and target points for prediction are denoted x_1 for eastings (or across the image from left to right) and x_2 for northings (or from bottom to top on the image). The pair $\{x_1, x_2\}$ are given the symbol $\mathbf{x}$ in vector notation. The variates or channels are denoted $z_1, z_2, \ldots$, and the measured values and recorded intensities of the pixels in the channels of the scanner are denoted $z(\mathbf{x}_i)$ for $i = 1, 2, \ldots$ for any one variate.

Screening

Few large files of data are free of mistakes caused by instrumental malfunction and human error. When you receive data, whether from remote scanners or from the field or laboratory, check for such mistakes.

Position. Examine the positions of the data in relation to the bounds of the region.

Plot them on a map, known as a 'posting' (**posting program**).

Do all the points lie within the region?

Are there sampling points or pixels in the sea when they should be on the land? If so why?

Have the coordinates been reversed inadvertently so that northings precede eastings, i.e. the x_2 precede x_1?

Do the pixels fill approximately the region?

Are the coordinates properly scaled?

Measurements. Screen the measured values, $z(\mathbf{x}_1), z(\mathbf{x}_2), \ldots$.

Pass them through a program that compares each value in the file against the minimum and maximum possible values of the scale and flags any that lie outside these bounds.

Print out the minimum and maximum for each z and check that they are sensible.

Histogram and distribution

For each variate compute a histogram and plot it, ensuring that all classes are of the same width (**histogram program**). You can also plot a cumulative distribution of z (**cumulative distribution program**). Examine it for outliers, i.e. individuals or small groups that are isolated from the main body of data. In addition, if you prefer box-plots, compute them of the zs to show outliers (**boxplot program**) or scatter diagrams (**scatter program**).

Outliers. If there are outliers identify their positions in the image or on the map.

Are they mistakes? If not what do they represent?

Are they part of the 'target population?

If not (e.g. houses when you are interested only in trees) then replace the recorded values by a symbol to indicate missing or not applicable.

The statistical treatment of outliers is a complex subject, and if you wish to retain outliers in your analysis then consult a statistician.

Frequency distribution. Study the shape of the histogram.

Has it more than one peak?

If so the scene or region almost certainly contains at least two distinct populations, e.g. land and water, farmland and forest.

Is the distribution symmetric?

If it is skewed is the longer tail towards the small values or towards the large?

Summary. Summarize the statistics for each variate by computing using **histogram program**

- the number of pixels or sampling points and the number of valid values,
- the minimum and maximum,
- the mean,
- the median,
- the variance,
- the standard deviation (the square root of the variance),
- the coefficient of variation (optional),
- the skewness (coefficient g_1), and
- the kurtosis (g_2, optional).

Normality and transformation

Geostatistical analysis is most efficient when done on variables that have normal, or Gaussian, distributions. Some analyses assume normality. You should therefore examine the form of the distribution of each z.

Symmetric histogram. The frequency distribution appears symmetric with a single central peak.

If so try fitting a normal curve to it.

If the fit 'looks good' then accept the variate as normal.

If not, in what way does it depart from normal, e.g. flat-topped, light in the tails? These may be matched with the coefficient of kurtosis, g_2. A flat topped distribution suggests that you have more than one population in the image or region—see multiple peaks above.

Skewed histogram. Asymmetry is the most common form of departure from normality, and in particular positive skewness (long upper tail, coefficient $g_1 > 0$).

In these circumstances the variance is likely to change from one part of the image or region to another, thereby violating one of the assumptions of stationarity on which analysis is usually based.

Consider transforming the recorded z to stabilize the variance.

Options are as follows.

- Skew positive, $0 < g_1 < 0.5$. Do not transform.
- Skew positive, $0.5 < g_1 < 1$. Consider transformation to square roots, i.e. $y = \sqrt{z}$.
- Skew positive, $g_1 > 1$. Transform. Try logarithmic transformation first, i.e. $y = \ln z$ or $y = \log_{10} z$. Examine the resultant distribution. If it is approximately normal then accept it. If the result is still skewed then try subtracting a positional constant, a, so that $y = \ln(z - a)$.

  A suitable value for a may be found by fitting the two- and three-parameter lognormal functions to the z.

Other transformations are available if these prove unsatisfactory.

Significance tests for normality are available.

Disregard them when analysing images! With many pixels you will almost surely discover that the distributions are 'significantly' non-normal. They can be helpful if you have only 100 or so measurements from ground survey.

Spatial distribution

Explore the spatial distribution of each z. Here you might need to treat image data differently from ground data.

Images. Map the distribution of pixel values using a congenial computer program that will show the individual pixels coloured, or shaded grey, on a scale according to recorded values, or transformed to $y = f(z)$.

Compute the *row and column means.* Alternatively, find the medians of each row and column.

Is there any trend in them?

Ground data. Make an isarithmic ('contour') map using a reputable program with a well-behaved algorithm for interpolation, such as inverse squared distance weighting or simple bilinear interpolation if the data are dense, and layer shading to indicate the magnitude of z, or $y = f(z)$.

Examine either kind of map for trends and patches.

- *Trend.* Is there any evident long-range trend over the scene or region, examine the data using **scatter and mean values program.**

  If so what is its form and principal direction?

- *Patches.* Are there patches?

  If so how big are they on average?

  Are they symmetric?

  If not in which direction are they elongated?

Long-range trend is incompatible with the assumptions of stationarity on which most geostatistical analysis is based. If the trend is strong then consider removing it by some kind of filter, such as a global trend surface, before proceeding further (**trend program**).
Alternatively, adopt a model for z that incorporates the non-stationary trend. This will take you into more advanced technique, and you should consult a specialist about it.

Spatial analysis—the variogram

The variogram summarizes the spatial distribution of z in the absence of trend. Three variograms are to be distinguished: the experimental variogram, the regional variogram, and the theoretical variogram.

The experimental variogram. This is the variogram that you compute from the data, $z(\mathbf{x}_i), i = 1, 2, \ldots$.

For images and ground data on regular rectangular grids compute it separately along the rows and columns of the grid, incrementing the lag by one pixel or sampling interval at a time, and along the principal diagonals of the grid at intervals of $\sqrt{2}$ pixels or sampling intervals **variogram 1 program.**

For irregularly scattered data, and provided you have sufficient (several hundred), compute a variogram in four or more directions by discretizing the lag by both distance and direction. Compute also the variogram ignoring direction, i.e. with lag in distance only.

Plot the results as variance against lag distance with a unique symbol for each direction.

Identify the main features, as follows:

- *Anisotropy.* Does the variogram have approximately the same form and values in all directions? If so then accept is as isotropic and compute an average experimental variogram over the four directions.

  If not then in what way do the directions differ?

  - *Different spatial scale.* This indicates geometric anisotropy, which might be removed by a simple transformation of the spatial coordinates.
  - *Different semivariances.* This indicate 'zonal' anisotropy—there is simply more variance in some directions than in others.
  - *Different form.* Look especially for contrasts between convex (decreasing gradient with increasing lag distance) and concave (increasing gradient). This suggests trend in the direction of increasing gradient, and it should be compared with the evidence from the exploratory analysis, above.

- *Bounds.* Does the variogram appear bounded, i.e. does the semivariance reach a maximum within the distance computed or appear as though it would reach a maximum if the lag distance was extended somewhat (bounded)?

  Alternatively, does it look as though it would increase without limit (unbounded)?

- *Nugget.* Does an imaginery line drawn through the experimental values when projected cut the ordinate at a positive value (not 0)? If so this intercept is known as the nugget variance.

The regional variogram. This is the variogram that you would compute of you had complete information in the region. It is approximated by the experimental variogram.

The theoretical variogram. This is the variogram of the process that you must imagine generated the scene or field of which the pixels or measured data are a sample.

To proceed further you must fit a mathematical function to the experimental variogram as a model or approximation to the theoretical variogram, see below.

Modelling the variogram

1. Match the form of the experimental variogram with those of the common simple valid models for variance in two dimensions. Choose several that appear to have the right form (**variogram 2 program**).
2. Fit each of these models in turn using a numerically sound and well-tried program by minimizing a weighted least squares criterion. Choose weights in proportion to the number of paired comparisons in the experimental values, and set approximate starting values for the non-linear parameters. Tabulate the residual sum of squares and residual mean square as criteria. You may use a more elaborate scheme of weights such as those of McBratney & Webster (1986) if you wish to model the variogram better near the ordinate.
3. Select the function for which the criteria are least. Plot the fitted function on the same pair of axes as the experimental semivariances.

   Does the fit appear good on the graph?

   If not inspect another.

   If none appear to fit well then consider fitting a more complex model by combining two or more simple models from the standard repertoire, and repeat the process.

   Akaike Information Criterion. In principle, you can always improve the fit of a model by making it more complex, i.e. by increasing the number of parameters in it.

   To compare functions with different numbers of parameters calculate the Akaike Information Criterion (AIC), and choose the model for which the AIC is least. This trades complexity against goodness of fit.

   The AIC is defined as

   $$\text{AIC} = -2\ln(\text{maximized likelihood}) + 2 \times (\text{numberof parameters}) \ .$$

   For any given experimental variogram it has a variable part:

   $$\widehat{A} = n \ln R + 2p \ ,$$

   where n is the number of experiemntal values, R is the residual sum of squares, and p is the number parameters.

   Least squares fitting minimizes R, but if it is diminished further only by increasing p (n is constant) then there is a penalty, which might be too big.

4. Check that the models that appear to fit accord with prior knowledge. If they do not then investigate further. You might need to shorten the interval between successive lags, narrow the angular discretization, or extend the maximum distance over which you compute the experimental variogram. You might need to try fitting other models.
5. Tabulate the parameters of the final best model and any others that are almost equally good. You will need the parameters for kriging (below).

Spatial estimation or prediction: kriging

The aim is to estimate or predict in a spatial sense the values of z at unsampled places, or 'targets', from the data. For images such targets are likely only where there are gaps in a scene. Ordinary kriging does smooth, however, and you might choose to use it to remove short-range noise in the image so that you can see a more general pattern (**kriging 1 and kriging to produce a contour map programs**). For ground surveys they are commonplace, and in this section ground survey is assumed. Further, ordinary kriging of z (or $y = \ln z$ for lognormal kriging) is likely to serve in 90% of cases, and only this is covered.

Essentials

You will need the original data and a legitimate model of the variogram.

Choices

Punctual or block kriging

The targets may be points, say $\mathbf{x}_0$, in which case the technique is punctual kriging. Alternatively they may be small blocks, B, which may be of any reasonable size and shape but are usually square; this is block kriging. The size of block should be determined by the application: what size of block does the user of the predictions want? It should not be determined by the data or the cosmetics of mapping—see below.

Number of data points

Ordinary kriging computes a weighted average of the data. The weights are determined by the configuration of the data in relation to the target in combination with the variogram model. They do not depend on $z(\mathbf{x}_i)$. Unless the model has a large proportion of nugget variance only the nearest few sampling

points carry appreciable weight; more distant points have negligible weight. So kriging is local.

Take the nearest 20 points to the target. If the data points are exceptionally unevenly scatterd then take the nearest two or three points in each octant around the target.

Form the kriging equations, and solve them to obtain the weights, the predicted values and the prediction variances (kriging variances).

If you are uncertain how many points to take then experiment with numbers between 4 and 40 and plot their positions in relation to the target and their weights. Do not be alarmed if some weights are negative, provided they are close to 0.

Transformation

For lognormal kriging the data must transformed to $y = \ln z$ or $y = \log_{10} z$, and the variogram model must be of y. If you want estimates to be of z then you must transform the predicted y back to z.

Kriging for mapping

Krige at the nodes of a fine square grid. Write the kriged estimates and kriging variances to a file. For an isarithmic display the interval of the grid should be chosen such that it is no more than 2 mm on the final hard copy. The optimality of kriging will not then be noticeably degraded by non-optimal interpolation in the graphics program.

The grid interval need not be related to the block size if you block krige. The blocks may overlap, or there may be gaps between them.

Mapping

Pass the file of kriged estimates and variances to a graphics program for the final display of the results as isarithms or small square cells. Choose colours or grey levels to represent the magnitude of the estimates and variances, as above.

Do not use graphics programs or geographic information systems for geostatistics unless

(a) you are in complete control, and

(b) you know that they do exactly what you want.

Genstat instructions for analysis

The analyses summarized above can be done in Genstat using the following commands. The data used as the example are of the exchangeable potassium (K) in the soil of a 80-ha farm (Broom's Barn) in Eastern England. The farm was sampled at 40-m intervals on a square grid, and bulked cores of soil to 20 cm were taken and analysed in the laboratory.

The measured variable (z) is here denoted by `z`, and the spatial coordinates ($\{x_1, x_2\}$) by `x` and `y`. Unless otherwise defined variables are vectors, or in the Genstat language, `variates`.

Summary statistics

Genstat enables you to obtain a statistical summary readily by means of standard functions:

```
calculate zbar=mean(z)
calculate zmed=median(z)
calculate zmax=maximum(z)
calculate zmin=minimum(z)
calculate zmed=median(z)
calculate zvar=var(z)
```

and

```
calculate zsdev=sqrt(zvar)
```

To obtain the skewness, β_1, and kurtosis, β_2, requires a little more elaboration:

```
calculate m3=(z-zvar)**3
calculate g1=mean(m3)
calculate beta1=g1/(zvar*zsdev)
```

and

```
calculate m4=(z-zvar)**4
calculate g2=mean(m4)
calculate beta2=g2/(zvar*zvar)
```

Histogram

The histogram of z is formed simply by the command

```
histogram [title=!t('Potassium')] z
```

for a device such as a line printer, and by

```
dhistogram [title=!t('Potassium')] z
```

for a high quality graph. The title within the square brackets is an option.

You are likely to want to specify the limits to the classes, or 'bins' in statistical jargon. So define a variate containing them:

```
variate [values=10,15...100] binlims
```

Then write

```
dhistogram [limits=binlims; title=!t('Potassium')] z
```

If you want to see what the frequency distribution looks like on the logarithmic scale then z is readily transformed by

```
calculate lz=log10(z)
```

and you can then replace z by lz in the above commands.

Cumulative distribution

The cumulative distribution of z can be formed by the following set of commands

```
calculate az=sort(z)
calculate cz=cum(az)
```

in which sort assembles the values in z in order from smallest to largest, and cum accumulates them. You will want to scale cz between 0 and 1, and to divide it therefore by the maximum value of z:

```
calculate mx=max(z)
```

and

```
calculate pz=cz/mx
```

To draw a graph of the cumulative distribution you can write the command

```
dgraph x=az; y=pz
```

Further, to show it on a Normal probability scale you can convert pz to 'Normal equivalent deviates', as follows

```
calculate nd=ned(pz)
dgraph x=az; y=nd
```

Posting

You can plot the data as a posting as follows. You should assemble the outline of the region of interest as pairs of coordinates in two variates, say ox and oy. Then you can write in Genstat

```
pen 1; linestyle=0; method=point; symbols=4
pen 2; linestyle=1; method=line; symbols=0; join=given
dgraph y=y,ox; x=x,oy; pen=1,2
```

The variogram

Experimental variogram

You will first want to compute (form) the experimental or sample variogram from your data. The is done using the command `fvariogram`. It is followed by options and parameters. Below is an example, in which the `fvariogram` command is preceeded by the declarations of two variates to hold the directions in which you want to compute the variogram and the angles subtended by the segments.

```
variate [nvalues=4] angles; values=!(0,45,90,135)
variate [nvalues=4] segs; values=!(45,45,45,45)
fvariogram [y=y; x=x; step=1; xmax=13; \
        directions=angles; segments=segs] z;\
        variogram=zgam; counts=zcounts; distances=midpts
```

The identifiers `zgam`, `zcounts` and `midpts` are matrices, and you will usually want the results as vectors (variates). This is readily done by

```
variate vgram [#angles], lag [#angles], count [#angles]
calculate vgram[ ]=zgam$[*;1...4]
calculate lag[ ]=midpts$[*;1...4]
calculate counts[ ]=zcounts$[*;1...4]
```

which you can then print and graph.

Fitting a model

Genstat has a procedure, `mvariogram`, for fitting several standard models to experimental variograms. You can call it by, for example,

```
mvariogram [model=spherical; print=model, summary, estimates; \ weighting=cour
zgam; counts=zcounts; distances=midpts
```

This will fit a spherical model to the experimental variogram in `zgam` and `midpts` with weights proportional to the counts in `zcounts`.
The models available are unbounded linear, bounded linear, circular, spherical, pentaspherical, exponential, gaussian, Whittle's (besselk1), and power.

The procedure makes use of the `fitnonlinear` command in Genstat, and you can write models of your own choosing with this command. For example, to compute fit a spherical model you can write

```
expression spherical
value=!e(c=((1.5*lag/a-0.5*(lag/a)**3*(lag.le.a)+(lag.gt.a)))
model [weights=counts] vrgam
rcycle a; initial=400
```

```
fitnonlinear [calculation=spherical] c
```

Notice that only the non-linear part of the model has to be described in the `value` command. The `rcycle` command sets an initial value for `a`. This is to ensure that the search for a solution starts in roughly the right place. Otherwise the program might never converge.

Kriging

The kriging facility in Genstat creates a grid of estimates (predictions) for mapping using the command `krige`, as follows.

```
krige [x=x; y=y; youter=!(1,31); xouter=!(1,18); \
yinner=!(5,20); xinner=!(3,15); block=!(0,0); radius=4.5; \
minpoints=7; maxpoints=20; interval=0.5] \
z; isotropy=isotropic; model=spherical; nugget=0.00476; \
sill=0.01528; range=10.8; predictions=krigest; variances=krigvar
```

Control

Remember to terminate each Genstat job with

```
stop
```

APPENDIX I

Report on Visit to Topographic Engineering Center November 1996

One purpose of this visit was to discuss the results of the work carried out in phase 2 of the project. All of the work has been completed, except for the final report, which we discussed in terms of what the emphasis should be. The main purpose was to try to set up an environment for some geostatistical analysis at TEC.

Day 1 In the morning I discussed the results of the kriging analysis and geostatistical simulation with Kevin Slocum, Paul Krause and Joni Jarrett. Kevin was keen to involve Jim Moëller and Ed Bösch in the data reconstruction analysis so that we could compare the results of different approaches. A meeting with Ed was set up as he works in a different building.

In the afternoon I transferred the mapping files from England. This took some time because the files are large and the transfer slow. William Diego set up a computer user name for me so that I could work independently.

Day 2 I started to work on genstat – a statistical package that TEC has bought for geostatistical analysis. I had a trial program that I knew worked well at Rothamsted Experimental Station where the package was written. I had error messages when I ran it. They suggested that certain statements needed to be changed. I tried several changes and still got error messages. I contacted Rothamsted and they proved again that it would run there. One problem was that communication with England was very slow. I am not sure where the e-mail messages were held up.

Jim Moëller loaded the new SPOT image that we are to work onfor the next project. It covers the fort at AP Hill, to the south of Fredericksburg. He and Kevin chose the two areas that we are to work on.

Ed Bösch took the small files of data from Benning which we had used for reconstructing back to the original 10 000 points using kriging and simulation. He was going to use his method of wavelets. We shall then compare the output from the three methods to see which gives the smallest error.

The method of wavlets appears to have some overlap with the kriging analysis that we used to filter the image. Jim Shine who was on the geostatistics course that I taught three years ago wanted to discuss suggestions for a PhD project. As he has a strong background in mathematics I suggested that he considers looking at the kind of overlap that exists between kriging analysis and wavelets, and how they could complement one another.

In the evening I read the genstat manual that was available (the main manual

had not been sent) and decided on some changes that I could try.

Day 3 Persevered with genstat and then discovered from Rothamsted that the version that they had been sent at TEC was 3.1 instead of 3.2. The latter is the one with the geostatistical routines. It meant that I could no longer proceed with that analysis. I tried to get some FORTRAN programs running to compute variograms, but to start with I had to use the FORTRAN compiler on the machine at Birmingham and do the analysis at TEC. This was difficult and eventually Bill Diego installed a compiler at TEC. He had to obtain some help to get it to work. In the meantimne I got some of the variogram model fitting routines to run using genstat. Until they have genstat 3.2 they can use a FORTRAN program to compute the variograms. Genstat is needed for the variogram modelling. I went through the model input and output in detail with Joni Jarrett.

Day 4 I analysed the new file of data for the wildlife area of AP Hill. This is a much smaller file than that for Fort Benning. I computed the variograms and the summary statistics. In the evening I fitted the models to these variograms because we ran out of time during the day and Kevin Slocum wanted the results before going into the field the next day. All three wavebands gave similar results concerning the spatial scale in the variation.

Day 5 On the final day of my visit Kevin Slocum, Jim Moëller, Paul Krause and I went to AP Hill to look at the vegetation and relief in the two parts of the image that we shall analyse. We walked through the forest to try to relate the change in ground cover with the results of the variogram analysis. The shorter scale of variation was about 200 m and this structure was evident, especially in the small patches of evergreen woodland among the predominantly deciduous forest. The latter represents the longer scale of variation. There is also considerable variation in the undergrowth and the relief.

While in the field we decided that the vegetation would be determined from transects on coloured photographs of the area. Ground information is vital for our analysis because without this cannot interpret our results precisely.

We then visited the ITAMS office to ask whether they had the photographs of the area that we want to work in. These were ordered and should be available to TEC shortly.

I have now set up the programs for TEC to compute the variograms with the appropriate data. There is also a new file for modelling variograms that has been expanded.

APPENDIX II

Kriging analysis: theory.

Once distinct spatial structures have been identified in the variogram we can try to separate their sources. This is effectively filtering. It enables us to isolate specifically local or regional sources of variation and, thereby, interpret them more easily than when they are combined. For complex imagery separating out the different sources of variation helps to show the spatial structure present at the different scales, and it reduces the noise (Bourgault & Marcotte, 1993; Goovaerts & Webster, 1994; Wen & Sinding-Larsen, 1997). It is done by what Matheron (1982) called 'kriging analysis', and it is based on the decomposition of $Z(\mathbf{x})$ into the sum of K separate orthogonal random functions $Z^k(\mathbf{x}), k = 1, 2, \ldots, K$, each with its basic variogram $g^k(\mathbf{h})$ of Equation (??):

$$Z(\mathbf{x}) = \sum_{k=1}^{K} Z^k(\mathbf{x}) + \mu \ , \tag{1}$$

such that

$$\mathrm{E}[Z^k(\mathbf{x})] = 0$$

and

$$\begin{aligned} \frac{1}{2}\mathrm{E}[\{Z^k(\mathbf{x}) - Z^k(\mathbf{x}+\mathbf{h})\}\{Z^{k'}(\mathbf{x}) - Z^{k'}(\mathbf{x}+\mathbf{h})\}] &= b^k g^k(\mathbf{h}) \quad \text{if} \quad k = k' \\ &= 0 \quad \text{otherwise} \ . \end{aligned} \tag{2}$$

Relation (2) expresses the mutual independence of the K random functions $Z^k(\mathbf{x})$. With this assumption, the nested model (??) is easily retrieved from relation (1). We assume that $Z(\mathbf{x})$ is second-order stationary, which accords with the results above (the variogram models have asymptotes). The $Z^k(\mathbf{x})$ are the spatial components, and they represent the behaviour of $Z(\mathbf{x})$ at the spatial scales defined by the distance parameters of the basic variogram functions, $g^k(\mathbf{h})$. In practice each spatial component is estimated as a linear combination of the observations $z(\mathbf{x}_i)$:

$$\widehat{Z}^k(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i^k z(\mathbf{x}_i) \ , \tag{3}$$

where n is the number of observations, $z(\mathbf{x}_i), i = 1, 2, \ldots, n$, used for the estimation, and the λ_i^k are the weights assigned to the observations.

The n weights are chosen to ensure that the estimate is unbiased and that the estimation variance is minimal. This leads to the kriging system:

$$\sum_{j=1}^{n} \lambda_j^k \gamma(\mathbf{x}_i, \mathbf{x}_j) - \psi = b^k g^k(\mathbf{x}_i, \mathbf{x}_0) \quad \text{for all} \quad i = 1, 2, \ldots, n \ ,$$

$$\sum_{j=1}^{n} \lambda_j^k = 0 \ . \tag{4}$$

This system is is solved to find the weights, λ_i^k, to insert into Equation (3). The quantity ψ is a Langrange multiplier. From Equation (1), $\mathrm{E}[Z^k(\mathbf{x})] = 0$, and so the weights sum to 0 to assure unbiasedness, not to 1 as in the ordinary kriging formulation, system (??).

To account for local non-stationarity the kriging is usually done in fairly small moving neighbourhoods centred on $\mathbf{x}_0$. Then it is necessary only that $Z(\mathbf{x})$ is locally stationary, or *quasi-stationary*, so that Equation (1) can be rewritten as

$$Z(\mathbf{x}) = \sum_{k=1}^{K} Z^k(\mathbf{x}) + \mu(\mathbf{x}) \ , \tag{5}$$

where $\mu(\mathbf{x})$ is a local mean which can be considered as a long-range spatial component.

Estimation of the long-range component, i.e. the local mean $\mu(\mathbf{x})$ and the spatial component with the largest range, can be affected by the size of the moving neighbourhood (Galli *et al.*, 1984). To estimate a spatial component with a given range, the diameter of the neighbourhood should be at least equal to that of the range. It frequently happens when the sampling density and the range are large that there are so many data within the chosen neighbourhood that only a small proportion of them is retained. Although modern computers can handle many data at a time, the number of data used must be limited to avoid instabilities when inverting very large covariance matrices. Further, even if all the data could be retained, only the nearest ones contribute to the estimate because they screen the more distant data. Consequently, the neighbourhood actually used is smaller than the neighbourhood specified, which means that the range of the estimated spatial component is smaller than the range apparent from the structural analysis. Galli *et al.* (1984) recognized this, and where data lie on a regular grid they proposed using only every second or every fourth point to cover a large enough area but still with sufficient data. Such selection is somewhat arbitrary, and we have adopted an alternative proposed by Jaquet (1989) which involves adding to the long-range spatial component the estimate of the local mean.

APPENDIX III

Genstat example programs for analysis

```
job 'Histogram and cumulative curve '
print ' Brooms Barn '
scalar top,bot,rex,ii
text [nval=2] title
text [nval=4] axlab;
text [nval=1] header; values=!T(' ')
text [nval=1] label
frame window=1,2,3,4; ylower=0.2,0.2,0.2,0.2; \
      yupper=0.65,0.65,0.65,0.65; \
      xlower=0.0,0.0,0.42,0.42; xupper=0.45,0.45,0.87,0.87
open 'bbhist.gra'; channel=4 ; filetype=graphics
device 4
open 'bball.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; format=!(0.0,6,5,7,10,*)] \
      xx, yy, k, logk
calculate k=k/(k.ge.-3)
calculate kbar=mean(k)
calculate kvar=var(k)
calculate ksd=sqrt(kvar)
calculate m3=(k-kbar)**3
calculate g3=mean(m3)
calculate beta1=g3/(kvar*ksd)
calculate m4=(k-kbar)**4
calculate g4=mean(m4)
calculate beta2=g4/(kvar*kvar)-3
print kbar, kvar, ksd, beta1, beta2; decimals=5
calculate logk=log10(k)
calculate klbar=mean(logk)
calculate klvar=var(logk)
```

```
calculate klsd=sqrt(klvar)
calculate ml3=(logk-klbar)**3
calculate gl3=mean(ml3)
calculate betal1=gl3/(klvar*klsd)
calculate ml4=(logk-klbar)**4
calculate gl4=mean(ml4)
calculate betal2=gl4/(klvar*klvar)-3
print klbar, klvar, klsd, betal1, betal2; decimals=5
calc ak=sort(k)
calc ck=cum(ak)
calc mk=max(ck)
calc pk=ck/mk
calc mx=max(k)
calc mn=min(k)
calc bark=mean(k)
calc midk=med(k)
calc alk=sort(logk)
calc clk=cum(alk)
calc mlk=max(clk)
calc plk=clk/mlk
calc mlx=max(logk)
calc mln=min(logk)
calc barlk=mean(logk)
calc midlk=med(logk)
variate [values=10,15...100] binlims
calc bot=0
calc bot=bot-0.1*abs(bot)
calc top=60
calc top=top+0.1*abs(top)
calc rex=105
pen 19; size=1.0
axes window=1; pen=19; \
     ylower=bot; yupper=top; xlower=0; xupper=rex; \
     xlp=*; xmp=*
pen 1; linestyle=1;colour=1;method=monotonic; \
     symbols=0; thickness=1.5; size=1.0
histogram [title=label] k
```

```
   dhistogram [limits=binlims; window=1; keywindow=0; title='Potassium']
k
   axes window=2; pen=19; \
        ylower=bot; yupper=top; xlower=5.2; xupper=rex; \
        xmarks=!(20,40...100); ymarks=!(0,1000); \
        xtitle='K'; ytitle='Frequency'
   pen 1; linestyle=1;colour=1;method=monotonic; \
        symbols=0; thickness=1.5; size=1.0
   dgraph [window=2;keywindow=0;screen=keep] 1000;1000
   variate [values=1.05,1.1...2.0] binlog
   calc bot=0
   calc bot=bot-0.1*abs(bot)
   calc top=60
   calc top=top+0.1*abs(top)
   calc rex=2.05
   pen 19; size=1.0
   axes window=3; pen=19; \
        ylower=bot; yupper=top; xlower=1; xupper=rex; \
        xlp=*; xmp=*
   pen 1; linestyle=1;colour=1;method=monotonic; \
        symbols=0; thickness=1.5; size=1.0
   histogram [title=label] logk
   dhistogram [limits=binlog; window=3; keywindow=0; \
        screen=keep] logk
   axes window=4; pen=19; ylower=bot; yupper=top; \
        xlower=1; xupper=rex; xmarks=!(1,1.2...2.0); \
        ymarks=!(0,1000); xtitle='log K'
   pen 1; linestyle=1;colour=1;method=monotonic; \
        symbols=0; thickness=1.5; size=1.0
   dgraph [window=4;keywindow=0;screen=keep] 1000;1000
   calc bot=0
   calc top=1
   pen 1; linestyle=1;colour=1;method=monotonic; \
        symbols=0; thickness=1.5; size=1.0
   help env,pic,cur
   axes window=2; pen=19; \
        ylower=bot; yupper=top; xlower=0; xupper=mx; \
```

```
      ymarks=!(0,0.2...1); xmarks=!(0,20...100); \
      xtitle='K'; ytitle='Cumulative sum'
help env,pic,cur
dgraph [window=2;keywindow=0] x=ak; y=pk; pen=1
axes window=4; pen=19; ylower=bot; yupper=top; \
      xlower=mln; xupper=mlx; xmarks=!(0,0.2...2); \
      ymarks=!(0,0.2...1); xtitle='Log K'
pen 1; linestyle=1;colour=1;method=monotonic; \
      symbols=0; thickness=1.5; size=1.0
dgraph [window=4;screen=keep;keywindow=0] x=alk; y=plk; pen=1
print ak,pk;decimals=4
pen 1; linestyle=1;colour=1;method=monotonic; \
      symbols=0; thickness=1.5; size=1.0
axes window=1; pen=19; ylower=bot; yupper=top; \
      xlower=0; xupper=mx; xmarks=!(0,20...100); \
      xlabels=!t('0','20','40','60','80','100'); \
      ymarks=!(0,0.2...1); \
      ylabels=!t('0','0.2','0.4','0.6','0.8','1.0'); \
      xtitle='K'; ytitle='Cumulative sum'
dgraph [window=1;keywindow=0] x=ak; y=pk; pen=1
stop
```

```
job 'posting'
"make a posting"
text [nval=2] title
text [nval=4] axlab;
text [nval=1] header; values=!T(' ')
text [nval=1] label
scalar top,bot,left,right
frame window=1; ylower=0.01; yupper=0.99; \
      xlower=0.02; xupper=0.7
open 'bbpost.gra'; channel=4 ; filetype=graphics
device 4
open 'bbpost.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; format=!(0.0,6.0,5.0,*)] bx, by
calc bot=min(by)
calc top=max(by)
calc left=min(bx)
calc right=max(bx)
graph [title=label; ytitle='Northing'; xtitle='Easting'; \
      ylower=bot;yupper=top;xlower=left;xupper=right; \
      nrows=37;ncolumns=61] y=by; x=bx
pen 19; size=1.0
axes window=1; style=box; pen=19; ytitle='Northing'; xtitle='Easting';
\
      ylower=bot; yupper=top; xlower=left; xupper=right; \
      xmarks=!(5,10,15,20); ymarks=!(5,10...35)
pen 1; linestyle=0; colour=1; method=point; \
      symbols=4; thickness=2.0; size=1.0;
dgraph [window=1; keywindow=0] y=by; x=bx; pen=1
stop
```

```
job 'Variogram'
print ' Brooms Barn '
scalar top,bot,rex,ii,pi
text [nval=435] points; values=!t(435(' '))
text [nval=2] title
text [nval=1] label
frame window=1,2,3,4,5,6; ylower=0.2,0.2,0.2,0.2,0.2,0.2; \
      yupper=0.65,0.65,0.65,0.65,0.85,0.85; \
      xlower=0.0,0.0,0.45,0.45,0,0.45; \
      xupper=0.45,0.45,0.9,0.9,0.45,0.9; \
      xmlower=0.09
open 'bbvar.gra'; channel=4 ; filetype=graphics
device 4
open 'bball.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; \
      format=!(0.0,6.0,5.0,7.1,*)] xx, yy, k
calculate k=k/(k.ge.0.00001)
calculate kbar=mean(k)
calculate kvar=var(k)
calculate ksd=sqrt(kvar)
print kbar, kvar, ksd
calculate logk=log10(k)
calculate klbar=mean(logk)
calculate klvar=var(logk)
calculate klsd=sqrt(klvar)
print klbar, klvar, klsd
variate [nvalues=4] angles; values=!(0,45,90,135)
variate [nvalues=4] segs; values=!(45,45,45,45)
fvariogram [y=yy; x=xx; step=1; xmax=13; \
      directions=angles; segments=segs] logk; \
      variogram=kgam; counts=kcounts; distances=midpts
variate vgram [#angles], lag [#angles], counts [#angles]
calculate vgram[ ]=kgam$[*;1...4]
```

```
calculate lag[ ]=midpts$[*;1...4]
calculate counts[ ]=kcounts$[*;1...4]
print lag[], vgram[], counts[]
mvariogram[print = monitor,model,summary,estimates; \
      model=spherical; constant=estimate; weight=cbyvar; \
      xupper=13] kgam; counts=kcounts; \
      distance=midpts
stop
```

```
job 'Variogram'
print ' Brooms Barn '
scalar top,bot,rex,ii,pi
text [nval=435] points; values=!t(435(' '))
text [nval=2] title
text [nval=1] label
frame window=1,2,3,4,5,6; ylower=0.2,0.2,0.2,0.2,0.2,0.2; \
      yupper=0.65,0.65,0.65,0.65,0.85,0.85; \
      xlower=0.0,0.0,0.45,0.45,0,0.45; \
      xupper=0.45,0.45,0.9,0.9,0.45,0.9; \
      xmlower=0.09
open 'bbvar1.gra'; channel=4 ; filetype=graphics
device 4
open 'bball.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; \
      format=!(0.0,6.0,5.0,7.1,*)] xx, yy, k
calculate k=k/(k.ge.0.00001)
calculate kbar=mean(k)
calculate kvar=var(k)
calculate ksd=sqrt(kvar)
print kbar, kvar, ksd
calculate logk=log10(k)
calculate klbar=mean(logk)
calculate klvar=var(logk)
calculate klsd=sqrt(klvar)
print klbar, klvar, klsd
fvariogram [y=yy; x=xx; step=1; xmax=13; \
      directions=0; segments=180] logk; \
      variogram=kgam; counts=kcounts; distances=midpts
" directions=angles; segments=segs] logk; \"
print midpts,kgam,kcounts
variate [nval=13] lag,gam,wt
calculate lag=midpts$[*;1]
```

```
calculate gam=kgam$[*;1]
calculate wt=kcounts$[*;1]
mvariogram[print = monitor,model,summary,estimates; \
      model=spherical; constant=estimate; weight=cbyvar; \
      xupper=13] kgam; counts=kcounts; \
      distance=midpts
scalar TWOOPI; VALUE=0.636619772
scalar INUG,ISILL,IRANGE,DSILL,DRANGE
scalar c00,c11,c22,a11,a22,a1,a2,top,bot,aa
calculate aa=1.4
calculate bot=0.0
calculate bot=bot*(bot.lt.0)
calculate top=0.035
calculate top=top*(top.gt.0)
scalar zbar,v,sd,kk, rex, ii
frame window=1,2,3,4; ylower=0.20,0.12,0.32,0.52; \
      yupper=0.75,0.47,0.87,0.87; \
      xlower=0.1,0.48,0.10,0.48; xupper=0.65,0.93,0.8,0.93
variate [nval=201] dpar1, dpar2
variate [nval=201] xlag; values=!(1...201)
variate [nval=201] yvar; values=!(1...201)
calculate rex=13.5
calculate xlag=(xlag-1.0)*rex/200
calculate wt=wt*(lag.LT.15)
calculate INUG=MIN(gam)
calculate ISILL=0.8*MAX(gam)
calculate IRANGE=MAX(lag)/4.0
calculate DRANGE=1.5*IRANGE
calculate DSILL=1.2*ISILL
calculate HSILL=0.4*ISILL
print 'Model:  unbounded linear'
model [WEIGHTS=wt] gam; RESIDUALS=R; FITTEDVALUES=F
fit lag
print '***** Spherical model *****'
print ' weights = npairs '
expression spherical; \
      value=!e(c = ((1.5*lag/a-0.5*(lag/a)**3)* \
```

```
      (lag.le.a)+(lag.gt.a)))
model [weights=wt; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 10
fitnonlinear [calculation=spherical] c
calc wt2=wt/(F*F)
expression spherical; \
      value=!e(c = ((1.5*lag/a-0.5*(lag/a)**3)* \
      (lag.le.a)+(lag.gt.a)))
model [weights=wt2; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 10
fitnonlinear [calculation=spherical] c
calc wt2=wt/(F*F)
rkeep estimates=kest
calc kest=reverse(kest)
calc kk=kest$[1]
calc dist=kest$[3]
calc a11=dist
calc c11=kest$[2]
calc wt2=wt/(F*F)
graph [nr=24;nc=65] F,gam;lag;meth=1,p
calc c00=kk
calc dpar1=xlag/a11
calc yvar=c00+c11*((1.5*dpar1-0.5*dpar1**3)*(dpar1.le.1)+(dpar1.gt.1))
graph [title=label; \
      ytitle='variance';xtitle='lag/40 m'; \
      ylower=bot;yupper=top;xlower=0;xupper=rex; \
      nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1.0
axes window=1; pen=19; \
      xtitle='Lag distance/40 m'; \
      ytitle='Variance'; \
      ylower=bot; yupper=top; xlower=0; xupper=rex; \
      xmarks=!(0,2...14); \
      ymarks=!(0,0.01...0.1); \
      ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
      symbols=0,-9; thickness=2.0; size=1
```

```
dgraph [window=1;keywindow=0;title='Spherical'] \
     y=yvar,gam; x=xlag,lag; pen=1,2
print '***** Negative exponential model *****'
print ' weights = pairs '
expression negex; \
     value=!e(c = (1.0-exp(-lag/a)))
model [weights=wt; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 4
fitnonlinear [calculation=negex] c
calc wt2=wt/(F*F)
expression negex; \
     value=!e(c = (1.0-exp(-lag/a)))
model [weights=wt2; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 4
fitnonlinear [calculation=negex] c
rkeep estimates=kest
calc kest=reverse(kest)
calc kk=kest$[1]
calc dist=kest$[3]
calc a11=dist
calc c11=kest$[2]
calc c00=kk
calc dpar1=xlag/a11
calc yvar=c00+c11*((1-exp(-dpar1)))
graph [title=label; \
     ytitle='Variance';xtitle='Lag/40 m'; \
     ylower=bot;yupper=top;xlower=0;xupper=rex; \
     nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1.0
axes window=1; pen=19; \
     xtitle='Lag distance/40 m'; \
     ytitle='Variance'; \
     ylower=bot; yupper=top; xlower=0; xupper=rex; \
     xmarks=!(0,2...20); ymarks=!(0,0.01...0.1); \
     ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
     symbols=0,-9; thickness=2.0; size=1.0
```

```
dgraph [window=1;keywindow=0;title='Exponential'] \
      y=yvar,gam; x=xlag,lag; pen=1,2
print '***** Power function *****'
print ' weights = npairs '
expression power; \
      value=!e(c = ( (lag**a)))
model [weights=wt; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 0.4
fitnonlinear [constant=omit; calculation=power] c
calc wt2=wt/(F*F)
expression power; \
      value=!e(c = ( (lag**a)))
model [weights=wt2; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 0.4
fitnonlinear [constant=omit; calculation=power] c
calc wt2=wt/(F*F)
rkeep estimates=kest
calc kest=reverse(kest)
calc kk=kest$[1]
calc dist=kest$[2]
calc a11=dist
calc c11=kest$[1]
calc wt2=wt/(F*F)
graph [nr=24;nc=65] F,gam;lag;meth=1,p
calc c00=0
calc dpar1=xlag**a
calc yvar=c00+c11*dpar1
graph [title=label; \
      ytitle='variance';xtitle='lag/40 m'; \
      ylower=bot;yupper=top;xlower=0;xupper=rex; \
      nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1.0
axes window=1; pen=19; \
      xtitle='Lag distance/40 m'; \
      ytitle='Variance'; \
      ylower=bot; yupper=top; xlower=0; xupper=rex; \
      xmarks=!(0,2...20); ymarks=!(0,0.01...0.1); \
```

```
      ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
      symbols=0,-9; thickness=2.0; size=1
dgraph [window=1;keywindow=0;title='Power function'] \
      y=yvar,gam; x=xlag,lag; pen=1,2
print '***** Circular model *****'
print ' weights = npairs '
expression circular; \
      value=!e(c = ( (1.0-TWOOPI*arccos((lag/a)*(lag.le.a)) + \
      TWOOPI*(lag/a)*sqrt(abs(1.0-(lag/a)**2)) \
      * (lag.le.a)+(lag.gt.a) ) ))
model [weights=wt; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 10
fitnonlinear [calculation=circular] c
calc wt2=wt/(F*F)
expression circular; \
      value=!e(c = ( (1.0-TWOOPI*arccos((lag/a)*(lag.le.a)) + \
      TWOOPI*(lag/a)*sqrt(abs(1.0-(lag/a)**2)) \
      * (lag.le.a)+(lag.gt.a) ) ))
model [weights=wt2; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 10
fitnonlinear [calculation=circular] c
calc wt2=wt/(F*F)
rkeep estimates=kest
calc kest=reverse(kest)
calc kk=kest$[1]
calc dist=kest$[3]
calc a11=dist
calc c11=kest$[2]
calc wt2=wt/(F*F)
graph [nr=24;nc=65] F,gam;lag;meth=1,p
calc c00=kk
calc dpar1=xlag/a11
calc yvar=c00+c11*((1.0-TWOOPI*arccos((dpar1.le.1)*dpar1) + \
      TWOOPI*dpar1*sqrt(abs(1.0-dpar1**2)*(dpar1.le.1)) \
      +(dpar1.gt.1)))
graph [title=label; \
```

```
      ytitle='variance';xtitle='lag/40 m'; \
      ylower=bot;yupper=top;xlower=0;xupper=rex; \
      nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1.0
axes window=1; pen=19; \
      xtitle='Lag distance/40 m'; \
      ytitle='Variance'; \
      ylower=bot; yupper=top; xlower=0; xupper=rex; \
      xmarks=!(0,2...50); \
      ymarks=!(0,0.01...0.1); \
      ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
      symbols=0,4; thickness=2.0; size=0.7
dgraph [window=1;keywindow=0;title='Circular'] \
      y=yvar,gam; x=xlag,lag; pen=1,2
print '***** PentaSpherical model *****'
print ' weights = npairs '
expression pspherical; \
value=!e(c = ((1.875*lag/a-1.25*(lag/a)**3+0.375*(lag/a)**5)*\
      (lag.le.a)+(lag.gt.a)))
model [weights=wt; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 10
fitnonlinear [calculation=pspherical] c
calc wt2=wt/(F*F)
expression pspherical; \
      value=!e(c = ((1.875*lag/a-1.25*(lag/a)**3+0.375*(lag/a)**5)* \
      (lag.le.a)+(lag.gt.a)))
model [weights=wt2; distribution=normal] gam; fittedvalues=F
rcycle a; initial = 10
fitnonlinear [calculation=pspherical] c
calc wt2=wt/(F*F)
rkeep estimates=kest
calc kest=reverse(kest)
calc kk=kest$[1]
calc dist=kest$[3]
calc a11=dist
calc c11=kest$[2]
```

```
calc wt2=wt/(F*F)
calc c00=kk
calc dpar1=xlag/a11
calc yvar=c00+c11*((1.875*dpar1-1.25*dpar1**3 \
     +0.375*dpar1**5)*(dpar1.le.1)+(dpar1.gt.1))
graph [title=label; \
     ytitle='variance';xtitle='lag/40 m';\
     ylower=bot;yupper=top;xlower=0;xupper=rex;\
     nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1.0
axes window=1; pen=19; \
     xtitle='Lag distance/40 m'; \
     ytitle='Variance'; \
     ylower=bot; yupper=top; xlower=0; xupper=rex; \
     xmarks=!(0,2...50); \
     ymarks=!(0,0.01...0.1); \
     ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
     symbols=0,4; thickness=2.0; size=0.7
dgraph [window=1;keywindow=0;title='Pentaspherical'] \
     y=yvar,gam; x=xlag,lag; pen=1,2
print '***** Whittles model (with Bessel function k1) *****'
print ' y = c0 + c*(1-x/a*k1(x/a))'
expression bessel[1]; !e(z=lag/a)
& bessel[2]; !e(bc=(z.gt.7.0))
& bessel[3]; !e(ba=(z.lt.1.0))
& bessel[4]; !e(bb=1.0-ba-bc)
& bessel[5]; !e(zz=z*z)
& bessel[6]; !e(ezz=exp(-z))
& bessel[7]; !e(bwk=log(z)-0.1159315)
& bessel[8]; !e(bwk=bwk*zz*(0.5+zz*(0.0625+zz*(0.00260417+zz/18432))))
& bessel[9]; !e(ba=ba*(1+bwk-zz*(0.25+zz* \
     (0.078125+zz*(0.0043403+0.00010625*zz)))))
& bessel[10]; !e(bb=bb*ezz*(12.1077-(11.0673+4.473799*z) \
     /(1+0.471589*z+0.0125524*zz)))
& bessel[11]; !e(bc=sqrt(z)*ezz*1.253314*bc)
& bessel[12]; !e(bc=bc*(1+((((0.2775764/z-0.1441956) \
```

```
      /z+0.102539)/z-0.1171875)/z+0.375)/z))
& bessel[13]; !e(bess=ba+bb+bc)
& bessel[14]; !e(c=(1.0-bess))
rcycle a; initial = 3; lower=0.0001
fitnonlinear [ calculation=bessel[]] c
calc wt2=wt/(F*F)
rkeep estimates=kest
calc kest=reverse(kest)
calc kk=kest$[1]
calc dist1=kest$[3]
calc k1=kest$[2]
print kk, k1,dist1
graph [nr=24;nc=65] F,gam;lag;meth=l,p
calc c00=kk
calc a11=dist1
calc c11=k1
calc dpar1=xlag/a11
calc dpar1=0.00001*(dpar1.lt.0.00001)+(dpar1.ge.0.00001)*dpar1
calc z1=dpar1
calc bz=(z1.gt.7.0)
calc bx=(z1.lt.1.0)
calc by=1.0-bx-bz
calc za=z1*z1
calc eza=exp(-z1)
calc bv=log(z1)-0.1159315
calc bv=bv*za*(0.5+za*(0.0625+za*(0.00260417+za/18432)))
calc bx=bx*(1+bv-za*(0.25+za* \
      (0.078125+za*(0.0043403+0.00010625*za))))
calc by=by*eza*(12.1077-(11.0673+4.473799*z1) \
      /(1+0.471589*z1+0.0125524*za))
calc bz=sqrt(z1)*eza*1.253314*bz
calc bz=bz*(1+((((0.2775764/z1-0.1441956) \
      /z1+0.102539)/z1-0.1171875)/z1+0.375)/z1)
calc bss=bx+by+bz
calc bss=(1.0-bss)
calc yvar=c00+c11*bss
graph [title=label; \
```

```
      ytitle='variance';xtitle='lag/40 m';\
      ylower=bot;yupper=top;xlower=0;xupper=rex;\
      nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1
axes window=1; pen=19; \
      ylower=bot; yupper=top; xlower=0; xupper=rex; \
      xmarks=!(0,2...20); \
      ymarks=!(0,0.01...0.1); \
      ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
      symbols=0,-9; thickness=2.3; size=1
dgraph [window=1; keywindow=0; title='Whittle model'] \
      y=yvar,gam; x=xlag,lag; pen=1,2
" Example for fitting a model through the origin "
print '***** Whittles model (with Bessel function k1) *****'
print ' y = c0 + c*(1-x/a*k1(x/a))'
expression bessel[1]; !e(z=lag/a)
& bessel[2]; !e(bc=(z.gt.7.0))
& bessel[3]; !e(ba=(z.lt.1.0))
& bessel[4]; !e(bb=1.0-ba-bc)
& bessel[5]; !e(zz=z*z)
& bessel[6]; !e(ezz=exp(-z))
& bessel[7]; !e(bwk=log(z)-0.1159315)
& bessel[8]; !e(bwk=bwk*zz*(0.5+zz*(0.0625+zz*(0.00260417+zz/18432))))
& bessel[9]; !e(ba=ba*(1+bwk-zz*(0.25+zz* \
      (0.078125+zz*(0.0043403+0.00010625*zz)))))
& bessel[10]; !e(bb=bb*ezz*(12.1077-(11.0673+4.473799*z) \
      /(1+0.471589*z+0.0125524*zz)))
& bessel[11]; !e(bc=sqrt(z)*ezz*1.253314*bc)
& bessel[12]; !e(bc=bc*(1+((((0.2775764/z-0.1441956) \
      /z+0.102539)/z-0.1171875)/z+0.375)/z))
& bessel[13]; !e(bess=ba+bb+bc)
& bessel[14]; !e(c=(1.0-bess))
rcycle a; initial = 3; lower=0.0001
fitnonlinear [constant=omit; calculation=bessel[]] c
calc wt2=wt/(F*F)
rkeep estimates=kest
```

```
calc kest=reverse(kest)
calc kk=0.0
calc dist1=kest$[2]
calc k1=kest$[1]
print kk, k1,dist1
graph [nr=24;nc=65] F,gam;lag;meth=1,p
calc c00=kk
calc a11=dist1
calc c11=k1
calc dpar1=xlag/a11
calc dpar1=0.00001*(dpar1.lt.0.00001)+(dpar1.ge.0.00001)*dpar1
calc z1=dpar1
calc bz=(z1.gt.7.0)
calc bx=(z1.lt.1.0)
calc by=1.0-bx-bz
calc za=z1*z1
calc eza=exp(-z1)
calc bv=log(z1)-0.1159315
calc bv=bv*za*(0.5+za*(0.0625+za*(0.00260417+za/18432)))
calc bx=bx*(1+bv-za*(0.25+za* \
     (0.078125+za*(0.0043403+0.00010625*za))))
calc by=by*eza*(12.1077-(11.0673+4.473799*z1) \
     /(1+0.471589*z1+0.0125524*za))
calc bz=sqrt(z1)*eza*1.253314*bz
calc bz=bz*(1+((((0.2775764/z1-0.1441956) \
     /z1+0.102539)/z1-0.1171875)/z1+0.375)/z1)
calc bss=bx+by+bz
calc bss=(1.0-bss)
calc yvar=c00+c11*bss
graph [title=label; \
     ytitle='variance';xtitle='lag/40 m';\
     ylower=bot;yupper=top;xlower=0;xupper=rex;\
     nrows=37;ncolumns=61] y=yvar,gam; x=xlag,lag; method=line,point
pen 19; size=1
axes window=1; pen=19; ylower=bot; yupper=top; xlower=0; xupper=rex;
\
     xmarks=!(0,2...20); ymarks=!(0,0.01...0.1); \
```

```
      ylabels=!t('0','0.01','0.02','0.03','0.04')
pen 1,2; linestyle=1,0;colour=1;method=monotonic,point; \
      symbols=0,-9; thickness=2.3; size=1
dgraph [window=1; keywindow=0; title='Whittle model'] \
      y=yvar,gam; x=xlag,lag; pen=1,2
stop
```

```
job 'Kriging'
print 'Brooms Barn'
scalar top,bot,rex,ii,pi
text [nval=435] points; values=!t(435(' '))
text [nval=2] title
text [nval=1] label
open 'bbkrige.gra'; channel=4 ; filetype=graphics
device 4
open 'bball.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; \
     format=!(0.0,6.0,5.0,7.1,*)] xx, yy, k
calculate k=k/(k.ge.0.00001)
calculate kbar=mean(k)
calculate kvar=var(k)
calculate ksd=sqrt(kvar)
calculate m3=(k-kbar)**3
calculate g3=mean(m3)
calculate beta1=g3/(kvar*ksd)
calculate m4=(k-kbar)**4
calculate g4=mean(m4)
calculate beta2=g4/(kvar*kvar)-3
print kbar, kvar, ksd, beta1, beta2
calculate logk=log10(k)
variate [nvalues=2] xlims; values=!(1,18)
variate [nvalues=2] ylimss; values=!(1,31)
variate [nvalues=2] xbounds; values=!(1,18)
variate [nvalues=2] ybounds; values=!(1,31)
variate [nvalues=2] xyblock; values=!(0.5,0.5)
krige [y=yy; x=xx; yinner=ybounds; xinner=xbounds; \
     block=xyblock; radius=4; minpoints=7; maxpoints=20; \
     interval=0.5] \
     logk; model=spherical; nugget=0.004828; sill=0.01531; \
     range=11.054; predictions=kest; variances=kkvar
```

```
print kest
print kkvar
scalar rows, columns
calc rows=61
calc columns=35
getattribute [attribute=rows,columns] kest; save=dim
calculate dim['rows']=reverse(dim['rows'])
calculate nrow=nvalues(dim['rows'])
matrix [rows=dim['rows']; columns=dim['columns']] kkest, kkkvar
calculate (kkest,kkkvar)$[nrow...1;*] = (kest,kkvar)$[1...nrow;*]
frame window=1,2; ylower=0,0; yupper=0.9,0.9; \
     xlower=0,0.65; xupper=0.65,0.99
axes 1; ylower=0.5; yupper=31.5; xlower=0.5; xupper=18.5
dcontour [interval=0.05; title='Log potassium'] kkest
dcontour [interval=0.0005; title='LogK variance'] kkkvar
stop
```

```
job 'Kriging'
print 'Brooms Barn'
scalar top,bot,rex,ii,pi
text [nval=435] points; values=!t(435(' '))
text [nval=2] title
text [nval=1] label
open 'bbkrige.gra'; channel=4 ; filetype=graphics
device 4
open 'bball.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; \
     format=!(0.0,6.0,5.0,7.1,*)] xx, yy, k
calculate k=k/(k.ge.0.00001)
calculate kbar=mean(k)
calculate kvar=var(k)
calculate ksd=sqrt(kvar)
calculate m3=(k-kbar)**3
calculate g3=mean(m3)
calculate beta1=g3/(kvar*ksd)
calculate m4=(k-kbar)**4
calculate g4=mean(m4)
calculate beta2=g4/(kvar*kvar)-3
print kbar, kvar, ksd, beta1, beta2
calculate logk=log10(k)
variate [nvalues=2] xlims; values=!(1,18)
variate [nvalues=2] ylimss; values=!(1,31)
variate [nvalues=2] xbounds; values=!(1,18)
variate [nvalues=2] ybounds; values=!(1,31)
variate [nvalues=2] xyblock; values=!(0.5,0.5)
krige [y=yy; x=xx; yinner=ybounds; xinner=xbounds; \
     block=xyblock; radius=4; minpoints=7; maxpoints=20; \
     interval=0.5] \
     logk; model=spherical; nugget=0.004828; sill=0.01531; \
     range=11.054; predictions=kest; variances=kkvar
```

```
print kest
print kkvar
scalar rows, columns
calc rows=61
calc columns=35
getattribute [attribute=rows,columns] kest; save=dim
calculate dim['rows']=reverse(dim['rows'])
calculate nrow=nvalues(dim['rows'])
matrix [rows=dim['rows']; columns=dim['columns']] kkest, kkkvar
calculate (kkest,kkkvar)$[nrow...1;*] = (kest,kkvar)$[1...nrow;*]
frame window=1,2; ylower=0,0; yupper=0.9,0.9; \
     lower=0,0.65; xupper=0.65,0.99
axes 1; ylower=0.5; yupper=31.5; xlower=0.5; xupper=18.5
dcontour [interval=0.05; title='Log potassium'] kkest
dcontour [interval=0.0005; title='LogK variance'] kkkvar
stop
```

```
job 'trend analysis'
text [nval=2] title
text [nval=1] header; values=!T(' ')
text [nval=1] label
frame window=1,2,3,4; ylower=0.25,0.05,0.5,0.5; \
      yupper=0.75,0.55,1.0,1.0; \
      xlower=0.0,0.45,0.0,0.45; xupper=0.5,0.95,0.5,0.95 ; \
      xmlower=0.08
open 'shine3.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; \
      format=!(0.0,-21,5.1,8.1,6.1,*)] x, y, z
calc xmin=min(x)
calc ymin=min(y)
calc zbar=mean(z)
calc zv=var(z)
calc sdz=sqrt(zv)
print xmin, ymin, zbar, zv, sdz; decimals=3
calc x=x-xmin
calc y=y-ymin
calc yy=y*y
calc xx=x*x
calc xy=x*y
model z; residuals=r; fittedvalues=f
terms x,y,xx,xy,yy,z
fit x,y
print x,y,z,r,f
stop
```

```
job 'trend analysis'
text [nval=2] title
text [nval=1] header; values=!T(' ')
text [nval=1] label
frame window=1,2,3,4; ylower=0.25,0.05,0.5,0.5; \
      yupper=0.75,0.55,1.0,1.0; \
      xlower=0.0,0.45,0.0,0.45; xupper=0.5,0.95,0.5,0.95 ; \
      xmlower=0.08
open 'shine3.dat'; channel=2
read [channel=2] title
print title
read [channel=2] label
print label
read [channel=2; setnvalues=y; \
      format=!(0.0,-21,5.1,8.1,-32,6.1,*)] x, y, z
calc xmin=min(x)
calc ymin=min(y)
calc zbar=mean(z)
calc zv=var(z)
calc sdz=sqrt(zv)
print xmin, ymin, zbar, zv, sdz; decimals=3
calc x=x-xmin
calc y=y-ymin
calc yy=y*y
calc xx=x*x
calc xy=x*y
model z; residuals=r; fittedvalues=f
terms x,y,xx,xy,yy,z
fit x,y,xx,yy,xy
print x,y,z,r,f
stop
```